5-1-2002

The Block-block Bootstrap: Improved Asymptotic Refinements

Donald W.K. Andrews

THE BLOCK-BLOCK BOOTSTRAP: IMPROVED ASYMPTOTIC REFINEMENTS

By

Donald W. K. Andrews

May 2002

COWLES FOUNDATION DISCUSSION PAPER NO. 1370

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS

YALE UNIVERSITY

Box 208281

New Haven, Connecticut 06520-8281

http://cowles.econ.yale.edu/

The Block-block Bootstrap: Improved Asymptotic Refinements

Donald W. K. Andrews[1]
Cowles Foundation for Research in Economics
Yale University

November 2001
Revised: April 2002

Abstract

The asymptotic refinements attributable to the block bootstrap for time series are not as large as those of the nonparametric iid bootstrap or the parametric bootstrap. One reason is that the independence between the blocks in the block bootstrap sample does not mimic the dependence structure of the original sample. This is the join-point problem.

In this paper, we propose a method of solving this problem. The idea is not to alter the block bootstrap. Instead, we alter the original sample statistics to which the block bootstrap is applied. We introduce *block* statistics that possess join-point features that are similar to those of the block bootstrap versions of these statistics. We refer to the application of the block bootstrap to block statistics as the block-block bootstrap. The asymptotic refinements of the block-block bootstrap are shown to be greater than those obtained with the block bootstrap and close to those obtained with the nonparametric iid bootstrap and parametric bootstrap.

Keywords: Asymptotics, block bootstrap, block statistics, Edgeworth expansion, extremum estimator, generalized method of moments estimator, maximum likelihood estimator, t statistic, test of over-identifying restrictions.

JEL Classification Numbers: C12, C13, C15.

1 Introduction

The principal theoretical attribute of bootstrap procedures is the asymptotic refinements they provide. That is, when properly applied, bootstrap tests have errors in null rejection probabilities that are of a smaller order of magnitude as the sample size, N, goes to infinity than those of standard asymptotic tests based on the delta method. Similarly, bootstrap confidence intervals (CIs) have coverage probability errors of a smaller order of magnitude than those of standard asymptotic CIs based on the delta method.

This paper is concerned with the magnitude of the asymptotic refinements of the block bootstrap for time series. These asymptotic refinements are not as large as those of the nonparametric iid bootstrap or the parametric bootstrap. For example, for iid observations, the error in rejection probability (ERP) of a one-sided bootstrap t test based on the nonparametric iid bootstrap is $O(N^{-1})$, e.g., see Hall (1992). In contrast, for stationary strong mixing observations, the ERP of a one-sided bootstrap t test based on non-overlapping or overlapping blocks is $O(N^{-1/2-\xi})$ for $0 < \xi < 1/4$, where ξ depends on the block length, see Andrews (2002), hereafter denoted A2002, and Zvingelis (2002). For the parametric bootstrap, the ERP of a one-sided bootstrap t test is essentially the same as that for the nonparametric iid bootstrap. This holds for iid observations as well as for stationary strong mixing Markov observations, see Andrews (2001b).

There are two reasons why the asymptotic refinements of the block bootstrap are less than those of the nonparametric iid bootstrap. The first is that the independence between the blocks in the block bootstrap sample does not mimic the dependence structure of the original sample. This is the *join-point* problem. The second reason is that the use of blocks of length greater than one increases the variability of various moments calculated under the block bootstrap distribution in comparison to their variability under the nonparametric iid bootstrap distribution. The reason is that the variability is determined by the amount of averaging that occurs over the blocks and longer blocks yields fewer blocks and, hence, fewer terms in the averages.

In this paper, we propose a method of solving the join-point problem. We do not alter the block bootstrap, because there does not seem to be a way to avoid its join-point feature. Rather, we alter the original sample statistics to which the block bootstrap is applied. We introduce *block* statistics (for the original sample) that have join-point features that resemble those of the block bootstrap versions of these statistics. We call the application of the block bootstrap to block statistics the *block-block* bootstrap.

The asymptotic refinements obtained by the block-block bootstrap are shown to be greater than those obtained by the standard block bootstrap. In fact, the block length can be chosen such that the magnitude of the asymptotic refinements of the block-block bootstrap is arbitrarily close to that obtained in the iid context using the nonparametric iid bootstrap. In practice, however, one would not expect the block-block bootstrap to perform as well as the nonparametric iid bootstrap for iid data. But, the asymptotic results suggest that it should outperform the block bootstrap in terms of ERPs and CI coverage probabilities.

A block statistic is constructed by taking a statistic that depends on one or more sample averages and replacing the sample averages by averages with some summands deleted. Let ℓ denote the block length to be used by the block bootstrap. We take ℓ such that $\ell = \ell_N \to \infty$ as $N \to \infty$. The join points of the block bootstrap sample are $\ell + 1$, $2\ell + 1, ...,$ $(b-1)\ell$, where b is the number of blocks and $N = b\ell$. We delete the $\lceil \pi\ell \rceil$ summands before each of the join points, where $\lceil \pi\ell \rceil$ denotes the smallest integer greater than or equal to $\pi\ell$, $\pi \in (0, 1)$, and $\pi = \pi_N \to 0$ and $\pi\ell - C\log(N) \to \infty$ as $N \to \infty$ for all constants $0 < C < \infty$. For example, $\pi \propto N^{-\delta\gamma}$ satisfies these conditions for any $0 < \delta < 1$. Note that π is the fraction of observations that are deleted from each block and from the whole sample.

For example, consider an estimator that minimizes a sample average of summands that depend on the observations and an unknown parameter θ, such as a quasi-maximum likelihood or least squares estimator. The corresponding block estimator minimizes the same sample average but with the summands described above deleted. A block t statistic for θ is based on a block estimator of θ normalized by a block standard deviation estimator.

Consider a sample average that appears in the definition of a block statistic. The last non-zero summand in one block is separated from the first summand in the next block by $\lceil \pi\ell \rceil$ time periods, where $\lceil \pi\ell \rceil \to \infty$ as $N \to \infty$. In consequence, for an asymptotically weakly dependent time series, such as a strong mixing process, the blocks are asymptotically independent. On the other hand, the blocks that appear in the bootstrap version of the block statistic are independent by construction.

Independence of the bootstrap blocks mimics the asymptotic independence of the original sample blocks sufficiently well that the join-point problem is solved. That is, join points do not affect the magnitude of the asymptotic refinements of the block-block bootstrap. See Section 2 for a detailed discussion of why this is true. Also, the join-point correction factors introduced in Hall and Horowitz (1996) and employed in A2002 for use with the block bootstrap are not needed with the block-block bootstrap.[2] Furthermore, in the case of an m-dependent process, the block length can be finite with the block-block bootstrap, whereas it must diverge to infinity with the standard block bootstrap.

Although the block-block bootstrap solves the join-point problem, the block-block bootstrap yields moments that are more variable than moments under the nonparametric iid bootstrap distribution—just as the standard block bootstrap does. In consequence, the asymptotic refinements obtained by the block-block bootstrap still depend on the block length. In particular, they are decreasing in the block length. Suppose $\ell \propto N^{\gamma}$ for some $0 < \gamma < 1$. In this paper, we show that the ERP of a one-sided bootstrap t test using the block-block bootstrap is $O(N^{-1/2-\xi})$ for all $\xi < 1/2 - \gamma$. In consequence, if γ is taken close to zero, the ERP is close to $O(N^{-1})$, which is the ERP of a one-sided nonparametric iid bootstrap t test.

In practice, one has to use a block length ℓ and a deletion fraction π that are large enough to accommodate the dependence in the data. Hence, one cannot just take γ arbitrarily close to zero. Thus, the above asymptotic result does not imply that one would expect the block-block bootstrap to work as well as the nonparametric iid bootstrap does with iid data. However, it does suggest that the block-block bootstrap should have smaller ERPs when $\gamma < 1/4$ than does the block bootstrap.

Block statistics have the same asymptotic efficiency as the standard statistics upon which they are based, because $\pi \to 0$ as $N \to \infty$. Hence, block-block bootstrap tests have the same asymptotic local power as standard asymptotic tests and as block bootstrap tests. Nevertheless, block statistics sacrifice some efficiency in finite samples because some observations are deleted. This is a drawback of the use of the block-block bootstrap.

A second drawback of the block-block bootstrap is that it requires the specification of the deletion fraction π, as well as the block length ℓ. The asymptotic results do not suggest a data-dependent method that is appropriate for choosing π and ℓ. In fact, the asymptotic refinements are maximized by taking γ arbitrarily close to zero, which is not a wise choice for finite samples, as stated above. (This implies that higher-order expansions are not helpful in choosing γ.) Also, the asymptotic refinements do not depend on π, provided π satisfies the conditions listed above, so they do not provide information regarding a good choice of π. It may be possible, however, to using higher-order expansions to determine a suitable choice of π for a given value of γ. This is beyond the scope of the present paper.

This paper presents the results of some Monte Carlo experiments that are designed to assess the finite sample performance of the block-block bootstrap. A dynamic regression model with regressors given by a constant, a lagged dependent variable, and three autoregressive variables is considered. Two-sided CIs for the coefficient on the lagged dependent variable are analyzed.

Standard delta method CIs are found to perform very poorly. For example, in the base case considered, which has a coefficient of .9 on the lagged dependent variable, a nominal 95% delta method CI has coverage probability of .759. Block and block-block bootstrap CIs are found to out-perform the delta method CI by a noticeable margin. For example, the usual nominal 95% symmetric non-overlapping block bootstrap CI with block lengths $\ell = 5$ and 10 has coverage probabilities of .922 and .915, respectively. The block-block bootstrap is found to improve upon the block bootstrap in most cases. For example, in the base case, with deletion fraction $\pi = .2$ and block lengths $\ell = 5$ and 10, the coverage probabilities of the symmetric block-block bootstrap are .928 and .938, respectively. The coverage probabilities of equal-tailed block and block-block bootstrap CIs are found to be noticeably worse than those of symmetric bootstrap CIs, but still noticeably better than those of delta method CIs.

In sum, the Monte Carlo results illustrate that the block-block bootstrap improves the finite sample performance of the block bootstrap in the dynamic regression models that are considered. The results also show that any of the bootstrap methods considered out-performs the delta method by a substantial margin.

We now discuss some alternative bootstraps for time series to the block bootstrap. One alternative is the parametric bootstrap for Markov time series. See Andrews (2001b) for an analysis of the higher-order improvements of this bootstrap. An obvious restriction of the parametric bootstrap is that it requires the existence of a parametric model or, at least, a conditional parametric model given some covariates.

Another alternative bootstrap procedure for Markov processes, that does not require a parametric model, is the Markov conditional bootstrap (MCB). Under some conditions, the MCB yields asymptotic refinements that exceed those of the block bootstrap, see Horowitz (2001). The MCB utilizes a nonparametric density estimator of the Markov transition density. This density has dimension equal to the product of the dimension of the observed data vector and the order of the Markov process plus one. For example, for a bivariate time series and a first-order Markov process, a four dimensional density needs to be estimated. Since nonparametric density estimators are subject to the curse of dimensionality, they are reliable only when the density has dimension less than or equal to three or, perhaps, four. In consequence, the range of application of the MCB is restricted to very low dimensional problems.

The tapered bootstrap of Paparoditis and Politis (2001, 2002) (PP) is another alternative to the block bootstrap. The tapered bootstrap is a variant of the block bootstrap in which the observations near the ends of the bootstrap blocks are down-weighted. PP shows that the tapered bootstrap is asymptotically correct to first order and that it reduces the asymptotic bias of the bootstrap variance estimator. PP does not address the issue of asymptotic refinements of the tapered bootstrap.

When the standard block bootstrap is applied to block statistics, as it is in this paper, the resulting bootstrap is a tapered bootstrap in which the tapering function is rectangular. Hence, the bootstrap procedure considered here is related to the tapered bootstrap of PP.[3] However, the key to obtaining the improved asymptotic refinements of the block-block bootstrap over the block bootstrap is that *both* the original sample statistic and the block bootstrap downweight observations near the end of the blocks. This is not considered in PP and it differentiates the approach taken in this paper from that of PP.

The discussion above indicates that the available alternatives to the block bootstrap for time series are useful, but are either only applicable in restrictive contexts or are not known to produce asymptotic refinements. In consequence, the problem addressed in this paper of how to increase the asymptotic refinements of the block bootstrap remains an important problem.

The results of this paper apply using the same assumptions and for the same cases as considered in A2002. In particular, two types of block bootstrap are considered—the non-overlapping block bootstrap, introduced by Carlstein (1986), and the overlapping block bootstrap, introduced by Künsch (1989). The results apply to extremum estimators, including quasi-maximum likelihood, least squares, and generalized method of moment (GMM) estimators. The results cover t statistics, Wald statistics, and J statistics based on the extremum estimators. One-sided, symmetric two-sided, and equal-tailed two-sided t tests and CIs are covered by the results. Tests of over-identifying restrictions are covered.

A key assumption made throughout the paper is that the estimator moment conditions are uncorrelated beyond some finite integer $\kappa \geq 0$, which implies that the covariance matrix of the estimator can be estimated using at most κ correlation estimates. This assumption is satisfied with $\kappa = 0$ in many time series models in which the estimator moment conditions form a martingale difference sequence due to optimizing behavior by economic agents, due to inheritance of this property from a regression error term, or due to the martingale difference property of the ML score function. It also holds with $0 < \kappa < \infty$ in many models with rational expectations and/or overlapping forecast errors, such as McCallum (1979), Hansen and Hodrick (1980), Brown and Maital (1981), and Hansen and Singleton (1982). For additional references, see Hansen and Singleton (1996). This assumption is also employed in A2002 and Hall and Horowitz (1996).

Some papers in the literature that do not impose the uncorrelatedness restriction beyond κ lags are Götze and Künsch (1996), Lahiri (1996), and Inoue and Shintani (2000). However, if the uncorrelatedness restriction does not hold and one employs a heteroskedasticity and autocorrelation consistent covariance matrix estimator, then the asymptotic refinements of the block bootstrap are smaller than otherwise and they depend on the choice of the smoothing parameter. The use of block statistics also may prove to have advantages in such cases. This is left to further research.

The proofs of the results in this paper make extensive use of the results of A2002. That paper, in turn, relies heavily on the methods used by Hall and Horowitz (1996), Bhattacharya and Ghosh (1978), Chanda and Ghosh (1979), Götze and Hipp (1983, 1994), and Bhattacharya (1987).

The paper A2002 considers the k-step block bootstrap as well as the standard block bootstrap. The asymptotic refinements established in this paper for the block-block bootstrap also hold for the k-step block bootstrap applied to block statistics provided the condition in A2002 on the magnitude of k is satisfied.

The remainder of the paper is organized as follows: Section 2 gives an overview of the problem considered in the paper and its solution based on block statistics. Section 3 defines the block extremum estimators. Section 4 defines the overlapping and non-overlapping block-block bootstraps. Section 5 states the assumptions. Section 6 establishes the asymptotic refinements of the block-block bootstrap. Section 7 reports some Monte Carlo results. An Appendix contains proofs of the results.

2 Overview of the Proposed Approach

In this section, we provide heuristic explanations of (i) the ERP of the standard one-sided asymptotic t test, (ii) the source of asymptotic refinements of the bootstrap, (iii) the join-point problem of the block bootstrap, and (iv) the improved asymptotic refinements attributable to the block-block bootstrap. Explanations of points (i) and (ii) are needed in order to exposit points (iii) and (iv).

2.1 ERP of the Standard Asymptotic t Test

We begin by discussing the ERP of the usual one-sided asymptotic t test. (The basic idea also applies to one-sided CIs and two-sided tests and CIs.) The observed sample is $\chi_N = \{X_i : i \le N\}$. We have an extremum estimator $\widehat{\theta}_N$ of an unknown parameter $\theta \in \Theta \subset R^{L_\theta}$. The estimator minimizes a sample average $\rho_N(\theta) = N^{-1}\sum_{i=1}^{N}\rho(X_i, \theta)$, where $\rho(\cdot, \cdot)$ is a known function. For example, $\widehat{\theta}_N$ could be a least squares or maximum likelihood estimator. (The results given below also cover the case where $\widehat{\theta}_N$ is a GMM estimator.)

We are interested in testing the null hypothesis $H_0 : \theta_r = \theta_{0,r}$ against the alternative hypothesis $H_1 : \theta_r > \theta_{0,r}$, where θ_r is the r-th element of θ. The t statistic for H_0 is

$$T_N = N^{1/2}(\widehat{\theta}_{N,r} - \theta_{0,r})/(\sigma_N)_{rr}^{1/2}, \tag{2.1}$$

where $\widehat{\theta}_{N,r}$ denotes the r-th element of $\widehat{\theta}_N$ and $(\sigma_N)_{rr}$ denotes an estimator of the asymptotic variance of $N^{1/2}(\widehat{\theta}_{N,r} - \theta_{0,r})$. The usual t test with asymptotic significance level α rejects H_0 if $T_N > z_\alpha$, where z_α is the $1 - \alpha$ quantile of a standard normal distribution.

Under suitable smoothness and moment conditions on $\rho(X_i, \theta)$, one can obtain an Edgeworth expansion for the distribution function (df) of T_N that holds under H_0. In particular, under H_0, we have

$$P(T_N \leq z) = \Phi(z) + N^{-1/2} q_N(z)\phi(z) + R_N(z), \tag{2.2}$$

where the remainder $R_N(z)$ satisfies $\sup_{z \in R} |R_N(z)| = O(N^{-1})$ as $N \to \infty$, the function $q_N(z)$ is $O(1)$ uniformly over z as $N \to \infty$, and $\phi(z)$ denotes the standard normal density function.

Taking $z = z_\alpha$ in the Edgeworth expansion of (2.2) gives

$$P(T_N > z_\alpha) = 1 - \Phi(z_a) - N^{-1/2} q_N(z_\alpha)\phi(z_\alpha) + O(N^{-1}) = 1 - \Phi(z_a) + O(N^{-1/2}). \tag{2.3}$$

Thus, the ERP under H_0 of the asymptotic t test is of order $O(N^{-1/2})$ as $N \to \infty$.

To understand the asymptotic refinements that the bootstrap can provide, the form of the function $q_N(z)$ in (2.2) is important. The function $q_N(z)$ is an even polynomial in z of order two. The coefficients of the polynomial in z are polynomials in moments of normalized partial derivatives of the criterion function $\rho_N(\theta)$ evaluated at θ_0, where θ_0 denotes the true value of θ under H_0. (The coefficients also depend on moments of normalized derivatives of random functions that arise in the standard error estimator $(\sigma_N)_{rr}^{1/2}$.) An example of such a moment is

$$E\left(N^{-1/2}\sum_{i=1}^{N}\frac{\partial}{\partial\theta_1}\rho(X_i, \theta_0)\right)^2, \tag{2.4}$$

where E denotes expectation under the distribution P that generates the sample χ_N and $E(\partial/\partial\theta_1)\rho(X_i, \theta_0) = 0$ by the asymptotic first-order conditions for the estimator.

If the observations are iid, the moment in (2.4) reduces to

$$E\left(\frac{\partial}{\partial\theta_1}\rho(X_i, \theta_0)\right)^2. \tag{2.5}$$

With dependent observations, the moment in (2.4) involves a double sum of correlations of random variables that depend on X_i and X_j for $i, j = 1, ..., N$.

2.2 Asymptotic Refinements of the Bootstrap

Next, we discuss the bootstrap and the asymptotic refinements that it achieves. The idea of the bootstrap is to estimate the joint distribution P of the observations by some estimator, say P^*, using the sample χ_N and to use this estimator to determine the distribution of T_N rather than to rely on the asymptotic distribution of T_N. For example, if the data are iid, one can take P^* to be the distribution of iid random vectors each of which has distribution given by the empirical distribution of the sample χ_N. This yields the *nonparametric iid* bootstrap. For dependent data, one can use the block bootstrap, which is described below.

Let $\chi_N^* = \{X_i^* : i \leq N\}$ be a sample of random vectors that are distributed according to P^* conditional on χ_N. Define $\rho_N^*(\theta)$ as $\rho_N(\theta)$ is defined, but with χ_N^* in place of χ_N; define θ_N^* to minimize $\rho_N^*(\theta)$ over Θ; define $(\sigma_N^*)_{rr}^{1/2}$ as $(\sigma_N)_{rr}^{1/2}$ is defined, but with χ_N^* in place of χ_N; and define the bootstrap t statistic $T_N^* = N^{1/2}(\theta_{N,r}^* - \widehat{\theta}_{N,r})/(\sigma_N^*)_{rr}^{1/2}$. The distribution of T_N^* under P^* mimics that of T_N under P, provided P^* is a suitable estimator of P. In consequence, the $1-\alpha$ quantile of T_N^*, denoted $z_{T,\alpha}^*$, can be used to approximate the $1-\alpha$ quantile of T_N. Hence, the bootstrap test of asymptotic significance level α rejects H_0 if $T_N > z_{T,\alpha}^*$.

Typically, the analytic calculation of $z_{T,\alpha}^*$ is intractable, but the simulation of samples χ_N^* with distribution P^* is easy and fast. In consequence, the bootstrap is carried out by (i) simulating a large number, B, of bootstrap samples $\chi_N^*(b) = \{X_i^*(b) : i \leq N\}$ for $b = 1, ..., B$, which are independent across samples (but not necessarily within samples) with each sample having distribution P^*; (ii) computing the B bootstrap criterion functions $\rho_N^*(\theta, b)$, estimators $\theta_N^*(b)$, and t statistics $T_N^*(b)$ for the bootstrap samples $\chi_N^*(b)$ for $b = 1, ..., B$; and (iii) approximating the population $1-\alpha$ quantile $z_{T,\alpha}^*$ of T_N^* by the sample $1-\alpha$ quantile $z_{T,\alpha}^*(B)$ of $\{T_N^*(b) : b = 1, ..., B\}$. As $B \to \infty$, $z_{T,\alpha}^*(B)$ converges in probability to $z_{T,\alpha}^*$, because a sample quantile of iid random variables converges in probability to the corresponding population quantile. Andrews and Buchinsky (2000) provide a three-step method for determining a value of B so that $z_{T,\alpha}^*(B)$ is close to $z_{T,\alpha}^*$ with high probability. Often, B needs to be in the range of 750–1000.

Under suitable conditions, one can obtain an Edgeworth expansion for the df of T_N^* that holds under H_0 conditional on χ_N. In particular, under H_0, we have

$$
\begin{aligned}
P^*(T_N^* \leq z) &= \Phi(z) + N^{-1/2} q_N^*(z)\phi(z) + R_N^*(z), \text{ where} \\
P(\sup_{z \in R} |R_N^*(z)| &> CN^{-1}) = O(N^{-1}) \text{ as } N \to \infty
\end{aligned}
\tag{2.6}
$$

for all constants $C > 0$. Note that the Edgeworth expansion in (2.6) is random because it depends on the original sample χ_N.

The function $q_N^*(z)$ in (2.6) is the same polynomial in z as $q_N(z)$, except that the moments that appear in the coefficients of the polynomial in z are taken with respect to P^* rather than P. Thus, the moment in (2.4) appears in the Edgeworth expansion of T_N^* with the expectation taken under P^* rather than P.

Using the Edgeworth expansions of (2.2) and (2.6) evaluated at $z = z_{T,\alpha}^*$ and the definition of $z^*_{T,\alpha}$, the ERP of the bootstrap t test under H_0 can be written as

$$\begin{aligned}
& P(T_N > z^*_{T,\alpha}) - \alpha \\
&= P(T_N > z^*_{T,\alpha}) - P^*(T^*_N > z^*_{T,\alpha}) \\
&= N^{-1/2}\left(q^*_N(z^*_{T,\alpha}) - q_N(z^*_{T,\alpha})\right)\phi(z^*_{T,\alpha}) + R^*_N(z^*_{T,\alpha}) - R_N(z^*_{T,\alpha}), \qquad (2.7)
\end{aligned}$$

where $z^*_{T,\alpha} \to_p z_\alpha$ as $N \to \infty$. The remainder terms are of smaller order of magnitude than the first term and, hence, can be ignored.

The ERP of the bootstrap t test depends on the magnitude of $q^*_N(z^*_{T,\alpha}) - q_N(z^*_{T,\alpha})$. This, in turn, depends on the magnitude of the difference between moments, such as that in (2.4), calculated under P^* and under P. Provided these differences are of order $o_p(N^{-\xi})$ for some $\xi \geq 0$, the ERP of the bootstrap can be shown to be $o(N^{-1/2-\xi})$, which is smaller than that of the standard asymptotic t test. The magnitude of these differences depends on how good an estimator P^* is of P, which depends on the type of bootstrap that is used.

2.3 Asymptotic Refinements of the Bootstrap for Iid Observations

If the data are iid and P^* is the nonparametric iid bootstrap distribution, then the expected value of a random variable under P^* equals the sample average of the random variable with respect to the sample χ_N. This holds because X^*_i has a discrete distribution with probability $1/N$ of equaling X_j for all $j = 1, ..., N$ under P^*. For example,

$$E^*\left(\frac{\partial}{\partial\theta_1}\rho(X^*_i, \theta_0)\right)^2 = N^{-1}\sum_{i=1}^{N}\left(\frac{\partial}{\partial\theta_1}\rho(X_i, \theta_0)\right)^2, \qquad (2.8)$$

where E^* denotes expectation with respect to P^*. By the central limit theorem,

$$E^*\left(\frac{\partial}{\partial\theta_1}\rho(X^*_i, \theta_0)\right)^2 - E\left(\frac{\partial}{\partial\theta_1}\rho(X_i, \theta_0)\right)^2 = O_p(N^{-1/2}). \qquad (2.9)$$

In consequence, the difference between the moment in (2.5) under P^* and under P is of order $N^{-1/2}$. Analogous results hold for other moments. Hence, the order of magnitude of $q^*_N(z^*_{T,\alpha}) - q_N(z^*_{T,\alpha})$ is $N^{-1/2}$. Using (2.7), this yields the order of magnitude of the ERP of the nonparametric iid bootstrap to be N^{-1}. Thus, the order of magnitude of the ERP of the nonparametric iid bootstrap t test, N^{-1}, is much smaller than that of the asymptotic t test, $N^{-1/2}$.

2.4 Asymptotic Refinements of the Block Bootstrap

We now discuss the *block* bootstrap for time series observations and, in particular, the magnitude of the reduction in ERP that can be obtained using the block bootstrap. We suppose that the observations $\{X_i : i \geq 1\}$ are a stationary strong mixing process with exponentially declining strong mixing numbers.

We consider two types of block bootstrap. One uses non-overlapping blocks and the other uses overlapping blocks. In either case, one breaks up the sample χ_N into blocks of length ℓ, where $\ell \propto N^{\gamma}$ for some $0 < \gamma < 1$. In the case of non-overlapping blocks, the first block is $(X_1, ..., X_\ell)$, the second block is $(X_{\ell+1}, ..., X_{2\ell})$, etc. There are b non-overlapping blocks, where $N = \ell b$. In the case of overlapping blocks, the first block is $(X_1, ..., X_\ell)$, the second block is $(X_2, ..., X_{\ell+1})$, etc. There are $N - \ell + 1$ overlapping blocks.

To obtain a block bootstrap sample, one draws b independent blocks by sampling with replacement from the non-overlapping or overlapping blocks which are based on the original sample. One lays the b randomly selected blocks end to end to form a sample of size $N = \ell b$. The distribution P^* for the block bootstrap is the distribution of the bootstrap sample obtained in the manner just described. Note that both the non-overlapping and the overlapping block bootstraps are obtained by drawing b independent blocks. The only difference is that one draws from b non-overlapping blocks in one case and from $N - \ell + 1$ overlapping blocks in the other.

The block bootstrap sample is $\{X_1^*, ..., X_\ell^*, X_{\ell+1}^*, ..., X_{2\ell}^*, ..., X_{(b-1)\ell+1}^*, ..., X_{b\ell}^*\}$. Within each randomly selected block, the dependence structure of the original sample is captured. But, the observation at the end of one block is independent of the next observation, which does not reflect the dependence pattern of the original data. In consequence, there is a problem with the joint points of the blocks in the bootstrap sample.

This has adverse consequences for the magnitude of the asymptotic refinements of the block bootstrap. It implies that there is a bias in the estimation of moments under P via moments under P^*. In consequence, the difference between moments in (2.4) under P^* and under P has order of magnitude no smaller than $N^{-\xi}$, where $\xi < \gamma$ and $\ell \propto N^{\gamma}$ for $0 < \gamma < 1$. Details are given below.

A second problem with the block bootstrap is that the variances of block bootstrap moments are larger than the variances of nonparametric iid bootstrap moments. This occurs because the use of b iid blocks of length ℓ, rather than N iid blocks of length one, leads to less averaging and, hence, larger variances. In consequence, the difference between moments in (2.4) under P^* and under P has order of magnitude no smaller than $N^{-\xi}$, where $\xi < 1/2 - \gamma$. Again, details are given below.

Given the two constraints on ξ, the upper bound on ξ is maximized when the block length ℓ is proportional to $N^{1/4}$, i.e., $\gamma = 1/4$. In this case, we obtain $\xi < 1/4$. Thus, the difference between moments in (2.4) under P^* and under P has order of magnitude no smaller than $N^{-1/4}$. Hence, the order of magnitude of $q_N^*(z_{T,\alpha}^*) - q_N(z_{T,\alpha}^*)$ is at least $N^{-1/4}$. Using (2.7), this yields the order of magnitude of the ERP to be at least $N^{-3/4}$. This shows that the order of magnitude of the ERP of the block bootstrap t test, which is at least $N^{-3/4}$, is smaller than that of the asymptotic t test, which is $N^{-1/2}$, as is desirable. But, it is larger than that of the nonparametric iid bootstrap t test, which is N^{-1}, which is not desirable. (Of course, the latter is not applicable when the data are dependent.)

We now show more explicitly why the restrictions $\xi < \gamma$ and $\xi < 1/2 - \gamma$ arise with the block bootstrap applied to standard statistics. We consider the non-overlapping block bootstrap. We compute the largest value ξ such that

$$E^*\left(N^{-1/2}\sum_{i=1}^{N}\frac{\partial}{\partial\theta_1}\rho(X_i^*,\theta_0)\right)^2 - E\left(N^{-1/2}\sum_{i=1}^{N}\frac{\partial}{\partial\theta_1}\rho(X_i,\theta_0)\right)^2 = o_p(N^{-\xi}). \quad (2.10)$$

Let b_j denote the set of ℓ indices of the j-th block for $j = 1, ..., b$. That is, $b_j = \{(j-1)\ell + 1, ..., j\ell\}$. Let b_j^* be the set of ℓ indices that correspond to the j-th randomly selected bootstrap block for $j = 1, ..., b$. Note that $\{b_j^* : j = 1, ..., b\}$ are iid each with a uniform distribution over $\{b_j : j = 1, ..., b\}$. Let

$$Y_j^* = \sum_{i\in b_j^*}\frac{\partial}{\partial\theta_1}\rho(X_i^*,\theta_0) \text{ and } Y_j = \sum_{i\in b_j}\frac{\partial}{\partial\theta_1}\rho(X_i,\theta_0). \quad (2.11)$$

We can rewrite (2.10) as

$$N^{-1}E^*\left(\sum_{j=1}^{b}Y_j^*\right)^2 - N^{-1}E\left(\sum_{j=1}^{b}Y_j\right)^2 = o_p(N^{-\xi}). \quad (2.12)$$

We have

$$E^*\left(\sum_{j=1}^{b}Y_j^*\right)^2 = \sum_{j=1}^{b}E^*(Y_j^*)^2 = b\left(\frac{1}{b}\sum_{j=1}^{b}Y_j^2\right), \quad (2.13)$$

where the first equality holds by independence across the bootstrap blocks and the second equality holds by identical distributions of the bootstrap blocks plus the fact that expectations under P^* for the non-overlapping block bootstrap are given by averages over the b blocks $b_1, ..., b_b$.

By (2.13), the bootstrap moment in (2.12) equals $N^{-1}\sum_{j=1}^{b}Y_j^2$. We can decompose the difference in (2.12) into the deviation of $N^{-1}\sum_{j=1}^{b}Y_j^2$ from its mean, $N^{-1}\sum_{j=1}^{b}EY_j^2$, and the difference between its mean and $N^{-1}E(\sum_{j=1}^{b}Y_j)^2$, which is a bias term. Using a strong mixing moment inequality derived in A2002 (see (9.51) and (9.53)), one can show that the deviation of $N^{-1}\sum_{j=1}^{b}Y_j^2$ from its mean is $o_p(N^{-\xi})$ provided $\xi < 1/2 - \gamma$.[4]

The bias term is

$$N^{-1}\sum_{j=1}^{b}EY_j^2 - N^{-1}E(\sum_{j=1}^{b}Y_j)^2 = -N^{-1}\sum_{\substack{j_1=1\\ j_1\neq j_2}}^{b}\sum_{j_2=1}^{b}EY_{j_1}Y_{j_2}. \quad (2.14)$$

The dominant terms in the double sum are the ones for which $|j_1 - j_2| = 1$, because if $|j_1 - j_2| \geq 2$, the summands of Y_{j_1} and Y_{j_2} differ by at least ℓ time periods, $\ell \to \infty$ as $N \to \infty$, and, hence, a strong mixing covariance inequality implies that these covariance terms are asymptotically negligible. The number of summands with $|j_1 - j_2| = 1$ is $2b-2$. We can show that $EY_{j_1}Y_{j_1+1} = O(1)$ using a covariance inequality for strong mixing random variables (because, for example, the first summand in Y_{j_1}

differs from the summands in Y_{j_1+1} by at least ℓ time periods). But, $EY_{j_1}Y_{j_1+1} \neq o(1)$, because the last observation in Y_{j_1} (or the next to last, etc.) is correlated with the first observation in Y_{j_1+1}.

Putting these results together, we find that the right-hand side (rhs) of (2.14) is

$$-N^{-1}((2b-2)O(1)+o(1)) = O(\ell^{-1}) = O(N^{-\gamma}) \tag{2.15}$$

using $N = b\ell$ and $\ell \propto N^{\gamma}$. The rhs is $o(N^{-\xi})$ only if $\xi < \gamma$. This demonstrates that the bias created by the join-point problem leads to the condition $\xi < \gamma$. In turn, this leads to smaller asymptotic refinements of the block bootstrap applied to standard statistics than those of the nonparametric iid bootstrap.

2.5 Asymptotic Refinements of the Block-block Bootstrap

The problem addressed in this paper is: how can one obtain bootstrap tests for time series data whose ERP is smaller than that of the block bootstrap and, hence, closer to that of the nonparametric iid bootstrap. To solve this problem, we propose altering the original sample t statistic, T_N, in such a way that it has block join points and its distribution is mimicked by the block bootstrap. We do so without affecting the asymptotic efficiency of the estimator $\widehat{\theta}_N$ that is used to construct the t statistic. In consequence, the proposed test has the same asymptotic local power as the asymptotic t test and the block bootstrap t test.

The idea is to delete some observations from the estimator criterion function $\rho_N(\theta) = N^{-1}\sum_{i=1}^{N}\rho(X_i,\theta)$ that occur just before the join points of the block bootstrap. The join points are the observations indexed by $\ell+1,\ 2\ell+1,\dots$. More specifically, before each join point, $\lceil \pi\ell \rceil$ observations are deleted, where π satisfies the conditions in the sixth paragraph of the Introduction. Let $\rho_\pi(X_i,\theta)$ denote $\rho(X_i,\theta)$ if the i-th observation is not deleted and 0 if the i-th observation is deleted. The resulting estimator criterion function is $\rho_{N,\pi}(\theta) = (N\tau)^{-1}\sum_{i=1}^{N}\rho_\pi(X_i,\theta)$, where $\tau = 1-\pi$. (We normalize the sum by $(N\tau)^{-1}$, rather than N^{-1}, because $N\tau$ is the number of non-zero terms in the sum.) We call $\rho_{N,\pi}(\theta)$ a *block* statistic and the estimator that minimizes it a *block* estimator.

When estimating the asymptotic standard deviation of $\widehat{\theta}_{N,r}$ that is used to construct the t statistic, we delete the same observations in the various sample averages that arise in the standard deviation estimator. The resulting t statistic is called a *block* t statistic.

The effect of deleting observations in $\rho_N(\theta)$ before the join points is that the last non-zero summand in each block of length ℓ is separated by length $\lceil \pi\ell \rceil$ from the first summand in the next block. Since $\pi\ell \to \infty$ as $N \to \infty$ and the observations are weakly asymptotically dependent (strong mixing), the summands in each block are asymptotically independent of the summands in the next block and every other block.

The bootstrap criterion function, call it $\rho^*_{N,\pi}(\theta)$, is the same as $\rho_{N,\pi}(\theta)$, but with the original sample replaced by the bootstrap sample. Note that $\rho^*_{N,\pi}(\theta)$ is comprised of independent blocks. This mimics the asymptotic independence of the blocks in

$\rho_{N,\pi}(\theta)$. The same is true for the bootstrap standard deviation estimator. Together, this eliminates the join-point problem because we can show that for all $\xi < 1/2 - \gamma$,

$$E^* \left(N^{-1/2} \sum_{i=1}^{N} \frac{\partial}{\partial \theta_1} \rho_\pi (X_i^*, \theta_0) \right)^2 - E \left(N^{-1/2} \sum_{i=1}^{N} \frac{\partial}{\partial \theta_1} \rho_\pi (X_i, \theta_0) \right)^2 = o_p(N^{-\xi}). \tag{2.16}$$

Analogous results hold for other moments. The key point is that the restriction $\xi < \gamma$, discussed above, is eliminated.

The block bootstrap applied to block statistics is called the block-block bootstrap. To see why the restriction $\xi < \gamma$ is eliminated with the block-block bootstrap, we look at the proof of (2.16). Let

$$Y_{\pi,j}^* = \sum_{i \in b_j^*} \frac{\partial}{\partial \theta_1} \rho_\pi (X_i^*, \theta_0) \text{ and } Y_{\pi,j} = \sum_{i \in b_j} \frac{\partial}{\partial \theta_1} \rho_\pi (X_i, \theta_0), \tag{2.17}$$

for $j = 1, ..., b$. By the same argument as in (2.13), the bootstrap moment in (2.16) equals $N^{-1} \sum_{j=1}^{b} Y_{\pi,j}^2$ and its deviation from its mean is $o_p(N^{-\xi})$ provided $\xi < 1/2 - \gamma$.

Similarly, the bias term is

$$N^{-1} \sum_{j=1}^{b} EY_{\pi,j}^2 - N^{-1} E(\sum_{j=1}^{b} Y_{\pi,j})^2 = -N^{-1} \sum_{\substack{j_1=1 \\ j_1 \neq j_2}}^{b} \sum_{j_2=1}^{b} EY_{\pi,j_1} Y_{\pi,j_2}. \tag{2.18}$$

At this point, the analysis differs from that of the previous section because even for $|j_1 - j_2| = 1$ the non-zero summands in Y_{π,j_1} and Y_{π,j_2} are separated by at least $\lceil \pi \ell \rceil$ time periods. In consequence, by a strong mixing covariance inequality of Davydov,

$$|EY_{\pi,j_1} Y_{\pi,j_2}| \leq K ||Y_{\pi,1}||_p^2 \alpha^r(\lceil \pi \ell \rceil), \tag{2.19}$$

for all $j_1, j_2 = 1, ..., b$, where K is a finite constant, $|| \cdot ||_p$ denotes the L_p norm, p and r are constants that satisfy $p, r \geq 1$ and $1/r + 2/p = 1$, and $\{\alpha(s) : s \geq 1\}$ are the strong mixing numbers of $\{X_i : i \geq 1\}$. This inequality can be used again to show that $||Y_{\pi,1}||_p = O(\ell)$ provided $E|(\partial / \partial \theta_1) \rho(X_i, \theta_0)|^p < \infty$. Hence, the rhs of (2.18) multiplied by N^ξ is

$$O(N^{\xi-1} b^2 \ell^2 \alpha^r(\lceil \pi \ell \rceil)) = O(N^{\xi+1} \alpha^r(\lceil \pi \ell \rceil)) = o(1), \tag{2.20}$$

where the last equality holds for all $\xi > 0$ because $\pi \ell - C \log(N) \to \infty$ as $N \to \infty$ for all constants $0 < C < \infty$ and the strong mixing numbers $\{\alpha(s) : s \geq 1\}$ decline exponentially fast. (The latter implies that $\alpha(s) \leq \beta_1 \exp(-\beta_2 s)$ for some $\beta_1, \beta_2 > 0$ and $N^{\xi+1} \alpha^r(\lceil \pi \ell \rceil)) \to 0$ iff $(\xi + 1) \log(N) + r \log(\alpha(\lceil \pi \ell \rceil)) \to -\infty$ iff $\lceil \pi \ell \rceil - (1/\beta_2) \log(\beta_1) - (\xi + 1)/(r\beta_2) \log(N) \to \infty$.)

We conclude that the bias term is $o(1)$ for all $\xi > 0$. These calculations show that the use of a block t statistic eliminates the bias problem associated with the block bootstrap. It does not, however, eliminate the variance problem. The restriction $\xi < 1/2 - \gamma$ still remains. This restriction is minimized by taking γ close to zero, which corresponds to taking short bootstrap blocks. Given the use of block statistics, this does not induce a bias problem.

Theoretically, one can choose γ arbitrarily close to zero and the ERP of the resulting block-block bootstrap t test has asymptotic order of magnitude arbitrarily close to N^{-1}, the same order as for the nonparametric iid bootstrap test. In practice, however, one needs to take $\gamma > 0$ sufficiently large that the dependence structure of the original sample is captured by the bootstrap blocks. Hence, one would not expect the block-block bootstrap to perform as well as the nonparametric iid bootstrap.

Nevertheless, the asymptotic results lead one to expect the block-block bootstrap to outperform the standard block bootstrap. That is, by considering block statistics when employing the block bootstrap, we expect to reduce the ERP of block bootstrap tests and the coverage probability errors of block bootstrap CIs.

In the following sections, we provide rigorous counterparts to the heuristic results discussed above.

3 Block Extremum Estimators and Tests

In this section, we define the block statistics that are considered in the paper. As much as possible, we use the same notation as A2002 and Hall and Horowitz (1996). The observations are $\{X_i : i = 1, ..., n\}$, where $X_i \in R^{L_x}$. The observations are assumed to be from a (strictly) stationary ergodic sequence of random vectors. We consider block versions of extremum estimators of an unknown parameter is $\theta \in \Theta \subset R^{L_\theta}$. The estimators we consider are either GMM estimators or estimators that minimize a sample average, which we call "minimum ρ estimators." Examples of minimum ρ estimators are maximum likelihood (ML), least squares (LS), and regression M estimators.

The GMM estimators that we consider are based on the moment conditions $Eg(X_i, \theta_0) = 0$, where $g(\cdot, \cdot)$ is a known L_g-valued function, X_i is as above, $\theta_0 \in \Theta \subset R^{L_\theta}$ is the true unknown parameter, and $L_g \geq L_\theta$. The minimum ρ estimators that we consider minimize a sample average of terms $\rho(X_i, \theta)$, where $\rho(\cdot, \cdot)$ is a known real function. Minimum ρ estimators can be written as GMM estimators with $g(X_i, \theta) = (\partial/\partial\theta)\rho(X_i, \theta)$.

We assume that the true moment vectors $\{g(X_i, \theta_0) : i \geq 1\}$ (for a GMM or minimum ρ estimator) are uncorrelated beyond lags of length κ for some $0 \leq \kappa < \infty$. That is, $Eg(X_i, \theta_0)g(X_{i+j}, \theta_0)' = 0$ for all $j > \kappa$. In consequence, the covariance matrix estimator and the asymptotically optimal weight matrix for the GMM estimator only depend on terms of the form $g(X_i, \theta)g(X_{i+j}, \theta)'$ for $0 \leq j \leq \kappa$. This means that the covariance matrix estimator and the weight matrix can be written as sample averages, which allows us to use the Edgeworth expansion results of Götze and Hipp (1983, 1994) for sample averages of stationary dependent random vectors, as in A2002 and Hall and Horowitz (1996). For this reason, we let

$$\widetilde{X}_i = (X_i', X_{i+1}', ..., X_{i+\kappa}')' \text{ for } i = 1, ..., n - \kappa. \tag{3.1}$$

All of the statistics considered below can be closely approximated by sample averages of functions of the random vectors $\widetilde{X}_i$ in the sample χ_N :

$$\chi_N = \{\widetilde{X}_i : i = 1, ..., N\}, \tag{3.2}$$

where $N = [(n-\kappa)/\ell]\ell$ for block bootstraps with block length ℓ and $[\cdot]$ denotes the integer part of $\cdot$. Thus, as in A2002, Hall and Horowitz (1996), and Götze and Künsch (1996), some observations $\widetilde{X}_i$ are dropped if $(n-\kappa)/\ell$ is not an integer to ensure that the sample size N is an integer multiple of the block length ℓ.[5]

Block statistics are based on sample averages of functions with certain summands deleted. The fraction of observations deleted is π, where π satisfies the conditions stated in the Introduction. As above, $\tau = 1 - \pi$ and ℓ is the block length. Given a function such as $g(X_i, \theta)$, we let $g_\pi(X_i, \theta)$ denote the function that is zero if the time subscript i corresponds to an observation that is one of the $\lceil \pi\ell \rceil$ observations before a join point and is $g(X_i, \theta)$ otherwise. Thus,

$$g_\pi(X_i, \theta) = \begin{cases} g(X_i, \theta) & \text{if } i \in [(j-1)\ell + 1, j\ell - \lceil \pi\ell \rceil] \text{ for some } j = 1, ..., b \\ 0 & \text{otherwise} \end{cases}. \tag{3.3}$$

We consider two forms of block GMM estimator. The first is a one-step block GMM estimator that utilizes an $L_g \times L_g$ non-random positive-definite symmetric weight matrix Ω. In practice, Ω is often taken to be the identity matrix I_{L_g}. The second is a two-step block GMM estimator that utilizes an asymptotically optimal weight matrix. It relies on a one-step block GMM estimator to define its weight matrix.

The one-step block GMM estimator, $\widehat{\theta}_N$, solves

$$\min_{\theta \in \Theta} J_{N,\pi}(\theta) = \left((N\tau)^{-1} \sum_{i=1}^{N} g_\pi(X_i, \theta) \right)' \Omega \left((N\tau)^{-1} \sum_{i=1}^{N} g_\pi(X_i, \theta) \right). \tag{3.4}$$

The two-step block GMM estimator which, for economy of notation, we also denote by $\widehat{\theta}_N$, solves

$$\begin{aligned} \min_{\theta \in \Theta} J_{N,\pi}(\theta, \widetilde{\theta}_N) &= \left((N\tau)^{-1} \sum_{i=1}^{N} g_\pi(X_i, \theta) \right)' \Omega_{N,\pi}(\widetilde{\theta}_N) \left((N\tau)^{-1} \sum_{i=1}^{N} g_\pi(X_i, \theta) \right), \text{ where} \\ \Omega_{N,\pi}(\theta) &= \overline{W}_{N,\pi}^{-1}(\theta), \\ \overline{W}_{N,\pi}(\theta) &= (N\tau)^{-1} \sum_{i=1}^{N} \left(g_\pi(X_i, \theta) g_\pi(X_i, \theta)' + \sum_{j=1}^{\kappa} H_\pi(X_i, X_{i+j}, \theta) \right), \\ H_\pi(X_i, X_{i+j}, \theta) &= g_\pi(X_i, \theta) g(X_{i+j}, \theta)' + g(X_{i+j}, \theta) g_\pi(X_i, \theta)', \end{aligned} \tag{3.5}$$

and $\widetilde{\theta}_N$ solves (3.4). By definition, $H_\pi(X_i, X_{i+j}, \theta)$ equals zero or not depending on the value of i, not $i + j$.

The block minimum ρ estimator, which we also denote by $\widehat{\theta}_N$, solves

$$\min_{\theta\in\Theta}\ (N\tau)^{-1}\sum_{i=1}^{N}\rho_\pi(X_i,\theta), \tag{3.6}$$

where $\rho_\pi(X_i,\theta)$ is defined analogously to $g_\pi(X_i,\theta)$ in (3.3) with $g(X_i,\theta)$ replaced by $\rho(X_i,\theta)$. For this estimator, we let $g_\pi(X_i,\theta)$ denote $(\partial/\partial\theta)\rho_\pi(X_i,\theta)$. Except for consistency properties, the block minimum ρ estimator can be analyzed simultaneously with the block GMM estimators. The reason is that with probability that goes to one (at an appropriate rate) the solution $\widehat{\theta}_N$ to the minimization problem (3.6) is an interior solution and, hence, is also a solution to the problem of minimizing a quadratic form in the first-order conditions from this problem with weight matrix given by the identity matrix, which is just the one-step block GMM criterion function.

The asymptotic covariance matrix, σ, of the block extremum estimator $\widehat{\theta}_N$ is

$$\sigma = \begin{cases} (D'\Omega D)^{-1}D'\Omega\Omega_0^{-1}\Omega D(D'\Omega D)^{-1} & \text{if } \widehat{\theta}_N \text{ solves (3.4)} \\ (D'\Omega_0 D)^{-1} & \text{if } \widehat{\theta}_N \text{ solves (3.5)} \\ D^{-1}\Omega_0^{-1}D^{-1} & \text{if } \widehat{\theta}_N \text{ solves (3.6), where} \end{cases}$$

$$\Omega_0 = \lim_{N\to\infty}(E\overline{W}_{N,\pi}(\theta_0))^{-1} \text{ and } D = E\frac{\partial}{\partial\theta'}g(X_i,\theta_0). \tag{3.7}$$

By stationarity, Ω_0 does not depend on π.

A consistent estimator of σ is

$$\sigma_{N,\pi} = \begin{cases} (D'_{N,\pi}\Omega D_{N,\pi})^{-1}D'_{N,\pi}\Omega\Omega_{N,\pi}^{-1}(\widehat{\theta}_N)\Omega D_{N,\pi} \\ \quad\times(D'_{N,\pi}\Omega D_{N,\pi})^{-1} & \text{if } \widehat{\theta}_N \text{ solves (3.4)} \\ (D'_{N,\pi}\Omega_{N,\pi}(\widehat{\theta}_N)D_{N,\pi})^{-1} & \text{if } \widehat{\theta}_N \text{ solves (3.5)} \\ D_{N,\pi}^{-1}\Omega_{N,\pi}^{-1}(\widehat{\theta}_N)D_{N,\pi}^{-1} & \text{if } \widehat{\theta}_N \text{ solves (3.6), where} \end{cases}$$

$$D_{N,\pi} = (N\tau)^{-1}\sum_{i=1}^{N}\frac{\partial}{\partial\theta'}g_\pi(X_i,\widehat{\theta}_N). \tag{3.8}$$

Let θ_r, $\theta_{0,r}$, and $\widehat{\theta}_{N,r}$ denote the r–th elements of θ, θ_0, and $\widehat{\theta}_N$ respectively. Let $(\sigma_{N,\pi})_{rr}$ denote the (r,r)–th element of $\sigma_{N,\pi}$. The block t statistic for testing the null hypothesis $H_0:\theta_r=\theta_{0,r}$ is

$$T_N = (N\tau)^{1/2}(\widehat{\theta}_{N,r}-\theta_{0,r})/(\sigma_{N,\pi})_{rr}^{1/2}. \tag{3.9}$$

Let $\eta(\theta)$ be an R^{L_η}–valued function (for some integer $L_\eta\geq 1$) that is continuously differentiable at θ_0. The block Wald statistic for testing $H_0:\eta(\theta_0)=0$ versus $H_1:\eta(\theta_0)\neq 0$ is

$$\mathcal{W}_N = (N\tau)\eta(\widehat{\theta}_N)'\left(\frac{\partial}{\partial\theta'}\eta(\widehat{\theta}_N)\sigma_{N,\pi}(\frac{\partial}{\partial\theta'}\eta(\widehat{\theta}_N))'\right)^{-1}\eta(\widehat{\theta}_N). \tag{3.10}$$

The block J statistic for testing over-identifying restrictions is

$$\begin{aligned} J_N &= K_{N,\pi}(\widehat{\theta}_N)' K_{N,\pi}(\widehat{\theta}_N), \text{ where} \\ K_N(\theta) &= \Omega_{N,\pi}^{1/2}(\theta)(N\tau)^{-1/2} \sum_{i=1}^{N} g_\pi(X_i, \theta) \end{aligned} \tag{3.11}$$

and $\widehat{\theta}_N$ is the block two-step GMM estimator. Under H_0, T_N has an asymptotic $N(0,1)$ distribution. If $L_g > L_\theta$ and the over-identifying restrictions hold, then J_N has an asymptotic chi-squared distribution with $L_g - L_\theta$ degrees of freedom. (This is not true if the one-step block GMM estimator is used to define the block J statistic.)

4 The Block-block Bootstrap

The observations to be bootstrapped are $\{\widetilde{X}_i : 1 \leq i \leq N\}$. As above, the block length ℓ satisfies $\ell \propto N^\gamma$ for some $0 < \gamma < 1$. (Note that one can take $\gamma = 0$ if the data are m-dependent.) We consider both non-overlapping and overlapping block bootstraps. For the non-overlapping block bootstrap, the first block is $\widetilde{X}_1, ..., \widetilde{X}_\ell$, the second block is $\widetilde{X}_{\ell+1}, ..., \widetilde{X}_{2\ell}$, etc. There are b different blocks, where $b\ell = N$. For the overlapping block bootstrap, the first block is $\widetilde{X}_1, ..., \widetilde{X}_\ell$, the second block is $\widetilde{X}_2, ..., \widetilde{X}_{\ell+1}$, etc. There are $N - \ell + 1$ different blocks.

The bootstrap is implemented by sampling b blocks randomly with replacement from either the b non-overlapping or the $N - \ell + 1$ overlapping blocks. Let $\widetilde{X}_1^*, ...,$ $\widetilde{X}_N^*$ denote the bootstrap sample obtained from this sampling scheme.

The bootstrap one-step block GMM estimator, θ_N^*, solves

$$\begin{aligned} \min_{\theta \in \Theta} J_{N,\pi}^*(\theta) &= \left((N\tau)^{-1} \sum_{i=1}^{N} g_\pi^*(X_i^*, \theta) \right)' \Omega \left((N\tau)^{-1} \sum_{i=1}^{N} g_\pi^*(X_i^*, \theta) \right), \text{ where} \\ g_\pi^*(X_i^*, \theta) &= g_\pi(X_i^*, \theta) - E^* g_\pi(X_i^*, \widehat{\theta}_N), \end{aligned} \tag{4.1}$$

X_i^* denotes the first element of $\widetilde{X}_i^*$, E^* denotes expectation with respect to the distribution of the bootstrap sample conditional on the original sample, and $g_\pi(X_i^*, \theta)$ is defined as $g_\pi(X_i, \theta)$ is defined in (3.3) but with X_i^* in place of X_i. For the non-overlapping and overlapping block bootstraps, respectively, we have:

$$\begin{aligned} (N\tau)^{-1} \sum_{i=1}^{N} E^* g_\pi(X_i^*, \theta) &= (N\tau)^{-1} \sum_{i=1}^{N} g_\pi(X_i, \theta) \text{ and} \\ (N\tau)^{-1} \sum_{i=1}^{N} E^* g_\pi(X_i^*, \theta) &= (N - \ell + 1)^{-1} \tau^{-1} \sum_{i=1}^{N} w(i, \ell, N) g_\pi(X_i, \theta), \text{ where} \\ w(i, \ell, N) &= \begin{cases} i/\ell & \text{if } i \in [1, \ell - 1] \\ 1 & \text{if } i \in [\ell, N - \ell + 1] \\ (N - i + 1)/\ell & \text{if } i \in [N - \ell + 2, N] \end{cases}. \end{aligned} \tag{4.2}$$

The bootstrap sample moments $(N\tau)^{-1}\sum_{i=1}^{N} g_\pi^*(X_i^*,\theta)$ in (4.1) are recentered (by subtracting off $E^* g_\pi(X_i^*,\widehat{\theta}_N)$) to ensure that their expectation $E^*(N\tau)^{-1}\sum_{i=1}^{N}$ $g_\pi^*(X_i^*,\theta)$ equals zero when $\theta=\widehat{\theta}_N$, which mimics the population moments $Eg_\pi(X_i,\theta)$, which equal zero when $\theta=\theta_0$. Note that recentering also appears in Shorack (1982), who considers bootstrapping robust regression estimators, as well as in Hall and Horowitz (1996) and A2002.

The bootstrap two-step block GMM estimator, also denoted by θ_N^*, solves

$$\min_{\theta\in\Theta} J_{N,\pi}^*(\theta,\widetilde{\theta}_N^*) = \left((N\tau)^{-1}\sum_{i=1}^{N} g_\pi^*(X_i^*,\theta)\right)' \Omega_{N,\pi}^*(\widetilde{\theta}_N^*) \left((N\tau)^{-1}\sum_{i=1}^{N} g_\pi^*(X_i^*,\theta)\right),$$

where

$$\begin{aligned}
\Omega_{N,\pi}^*(\theta) &= \overline{W}_{N,\pi}^*(\theta)^{-1}, \\
\overline{W}_{N,\pi}^*(\theta) &= (N\tau)^{-1}\sum_{i=1}^{N}\left(g_\pi^*(X_i^*,\theta)g_\pi^*(X_i^*,\theta)' + \sum_{j=1}^{\kappa} H_\pi^*(X_i^*,X_{i,i+j}^*,\theta)\right), \\
H_\pi^*(X_i^*,X_{i,i+j}^*,\theta) &= g_\pi^*(X_i^*,\theta)g^*(X_{i,i+j}^*,\theta)' + g^*(X_{i,i+j}^*,\theta)g_\pi^*(X_i^*,\theta)', \qquad (4.3)
\end{aligned}$$

$\widetilde{\theta}_N^*$ denotes the bootstrap one-step block GMM estimator that solves (4.1), and $X_{i,i+j}^*$ denotes the $(j+1)$-st element of $\widetilde{X}_i^*$ for $j=1,...,\kappa$.

The bootstrap block minimum ρ estimator, also denoted by θ_N^*, solves

$$\min_{\theta\in\Theta}\ (N\tau)^{-1}\sum_{i=1}^{N}(\rho_\pi(X_i^*,\theta) - E^* g_\pi(X_i^*,\widehat{\theta}_N)'\theta), \qquad (4.4)$$

where $g_\pi(\cdot,\theta)=(\partial/\partial\theta)\rho_\pi(\cdot,\theta)$. For the non-overlapping block bootstrap, the term $(N\tau)^{-1}\sum_{i=1}^{N} E^* g_\pi(X_i^*,\widehat{\theta}_N)'\theta$ is zero, because $(N\tau)^{-1}\sum_{i=1}^{N} E^* g_\pi(X_i^*,\widehat{\theta}_N) = (N\tau)^{-1}\sum_{i=1}^{N} g_\pi(X_i,\widehat{\theta}_N)=0$, where the second equality holds by the first-order conditions for $\widehat{\theta}_N$ using the fact that the dimensions of $g_\pi(\cdot,\cdot)$ and θ are equal. For the overlapping block bootstrap, $(N\tau)^{-1}\sum_{i=1}^{N} E^* g_\pi(X_i^*,\widehat{\theta}_N) \neq (N\tau)^{-1}\sum_{i=1}^{N} g_\pi(X_i,\widehat{\theta}_N)$ $=0$ and the extra term in (4.4) is non-zero. In this case, the term $(N\tau)^{-1}\sum_{i=1}^{N}$ $E^* g_\pi(X_i^*,\widehat{\theta}_N)'\theta$ properly recenters the block minimum ρ bootstrap criterion function. It yields bootstrap population first-order conditions that equal zero at $\widehat{\theta}_N$, as desired. That is, $E^*(\partial/\partial\theta)((N\tau)^{-1}\sum_{i=1}^{N}(\rho_\pi(X_i^*,\theta) - E^* g_\pi(X_i^*,\widehat{\theta}_N)'\theta)) = E^*(N\tau)^{-1}\sum_{i=1}^{N}$ $g_\pi^*(X_i^*,\theta)=0$ when $\theta=\widehat{\theta}_N$. With this recentering, the first-order conditions for θ_N^* are $(N\tau)^{-1}\sum_{i=1}^{N} g_\pi^*(X_i^*,\theta_N^*)=0$, rather than $(N\tau)^{-1}\sum_{i=1}^{N} g_\pi(X_i^*,\theta_N^*)=0$, which means that θ_N^* minimizes the one-step block GMM bootstrap criterion function $J_{N,\pi}^*(\theta)$ with $g_\pi(\cdot,\theta)=(\partial/\partial\theta)\rho_\pi(\cdot,\theta)$ and arbitrary positive definite weight matrix Ω.

The bootstrap block covariance matrix estimator is

$$\sigma_{N,\pi}^* = \sigma_{N,\pi}^*(\theta_N^*),\ \text{where}$$

$$\sigma^*_{N,\pi}(\theta) = \begin{cases} (D^*_{N,\pi}(\theta)'\Omega D^*_{N,\pi}(\theta))^{-1} D^*_{N,\pi}(\theta)\Omega\Omega^*_{N,\pi}(\theta)^{-1} & \\ \times\Omega D^*_{N,\pi}(\theta)(D^*_{N,\pi}(\theta)'\Omega D^*_{N,\pi}(\theta))^{-1} & \text{if } \widehat{\theta}_N \text{ solves (3.4)} \\ (D^*_{N,\pi}(\theta)'\Omega^*_{N,\pi}(\theta)D^*_{N,\pi}(\theta))^{-1} & \text{if } \widehat{\theta}_N \text{ solves (3.5)} \\ D^*_{N,\pi}(\theta)^{-1}\Omega^*_{N,\pi}(\theta)^{-1}D^*_{N,\pi}(\theta)^{-1} & \text{if } \widehat{\theta}_N \text{ solves (3.6) and} \end{cases}$$

$$D^*_{N,\pi}(\theta) = (N\tau)^{-1}\sum_{i=1}^{N}\frac{\partial}{\partial\theta'}g_\pi(X^*_i,\theta). \tag{4.5}$$

The bootstrap block t, Wald, and J statistics are

$$\begin{aligned} T^*_N &= (N\tau)^{1/2}(\theta^*_{N,r} - \widehat{\theta}_{N,r})/\sigma^*_{N,\pi}(\theta^*_N)^{1/2}_{rr}, \\ \mathcal{W}^*_N &= H^*_{N,\pi}(\theta^*_N)'H^*_{N,\pi}(\theta^*_N), \text{ and} \\ J^*_N &= K^*_{N,\pi}(\theta^*_N)'K^*_{N,\pi}(\theta^*_N), \text{ where} \\ H^*_{N,\pi}(\theta) &= \left((\frac{\partial}{\partial\theta'}\eta(\theta))\sigma^*_{N,\pi}(\theta)(\frac{\partial}{\partial\theta'}\eta(\theta))'\right)^{-1/2}(N\tau)^{1/2}(\eta(\theta) - \eta(\widehat{\theta}_N)), \\ K^*_{N,\pi}(\theta) &= \Omega^*_N(\theta)^{1/2}(N\tau)^{-1/2}\sum_{i=1}^{N}g^*_\pi(X^*_i,\theta), \end{aligned} \tag{4.6}$$

$\theta^*_{N,r}$ denotes the r-th element of $\theta^*_N$and $\sigma^*_{N,\pi}(\theta^*_N)_{rr}$ denotes the (r,r)–th element of $\sigma^*_{N,\pi}(\theta^*_N)$. Note that the bootstrap block t, Wald, and J statistics are *not* defined using correction factors, in contrast to the test statistics considered in Hall and Horowitz (1996) and A2002. Because of the block nature of the statistics, we do not have to correct for the fact that the bootstrap blocks are independent.

Let $z^*_{|T|,\alpha}$, $z^*_{T,\alpha}$, $z^*_{\mathcal{W},\alpha}$, and $z^*_{J,\alpha}$ denote the $1-\alpha$ quantiles of $|T^*_N|$, T^*_N, $\mathcal{W}^*_N$, and J^*_N respectively. To be precise, since the distributions of $|T^*_N|$ etc. are discrete, we define $z^*_{|T|,\alpha}$ to be a value that minimizes $|P^*(|T^*_N| \leq z) - (1-\alpha)|$ over $z \in R$. The precise definitions of $z^*_{T,\alpha}$, $z^*_{\mathcal{W},\alpha}$, and $z^*_{J,\alpha}$ are analogous.

Each of the following tests is of asymptotic significance level α. The symmetric two-sided block-block bootstrap t test of $H_0 : \theta_r = \theta_{0,r}$ versus $H_1 : \theta_r \neq \theta_{0,r}$ rejects H_0 if $|T_N| > z^*_{|T|,\alpha}$. The equal-tailed two-sided block-block bootstrap t test for the same hypotheses rejects H_0 if $T_N < z^*_{T,1-\alpha/2}$ or $T_N > z^*_{T,\alpha/2}$. The one-sided block-block bootstrap t test of $H_0 : \theta_r \leq \theta_{0,r}$ versus $H_1 : \theta_r > \theta_{0,r}$ rejects H_0 if $T_N > z^*_{T,\alpha}$. The block-block bootstrap Wald test of $H_0 : \eta(\theta_0) = 0$ versus $H_1 : \eta(\theta_0) \neq 0$ rejects the null hypothesis if $\mathcal{W}_N > z^*_{\mathcal{W},\alpha}$. The block-block bootstrap J test of over-identifying restrictions rejects the null if $J_N > z^*_{J,\alpha}$.

Each of the following CIs is of asymptotic confidence level $100(1-\alpha)\%$. The symmetric two-sided block-block bootstrap CI for $\theta_{0,r}$ is $[\widehat{\theta}_{N,r} - z^*_{|T|,\alpha}(\sigma_N)^{1/2}_{rr}/N^{1/2}$, $\widehat{\theta}_{N,r} + z^*_{|T|,\alpha}(\sigma_N)^{1/2}_{rr}\ /N^{1/2}]$. The equal-tailed two-sided block-block bootstrap CI for $\theta_{0,r}$ is $[\widehat{\theta}_{N,r} - z^*_{T,\alpha/2}(\sigma_N)^{1/2}_{rr}/N^{1/2}$, $\widehat{\theta}_{N,r} + z^*_{T,1-\alpha/2}(\sigma_N)^{1/2}_{rr}/N^{1/2}]$. The upper one-sided block-block bootstrap CI for $\theta_{0,r}$ is $[\widehat{\theta}_{N,r} - z^*_{T,\alpha}(\sigma_N)^{1/2}_{rr}/N^{1/2}$, $\infty)$. The block-block Wald-based bootstrap confidence region for $\eta(\theta_0)$ is $\{\eta \in R^{L_\eta} : N(\eta(\widehat{\theta}_N) - \eta)'((\partial\eta(\widehat{\theta}_N)/\partial\theta')\sigma_{N,\pi}(\partial\eta(\widehat{\theta}_N)/\partial\theta')')^{-1}(\eta(\widehat{\theta}_N) - \eta) \leq z^*_{\mathcal{W},\alpha}\}$.

5 Assumptions

We now introduce the assumptions. They are essentially the same as those of A2002 and are similar to those of Hall and Horowitz (1996).

Let $f(\widetilde{X}_i, \theta)$ denote the vector containing the unique components of $g(X_i, \theta)$ and $g(X_i, \theta)g(X_{i+j}, \theta)'$ for $j = 0, ..., \kappa$, and their derivatives through order $d_1 \geq 3$ with respect to θ. Let $(\partial^j / \partial\theta^j)g(X_i, \theta)$ and $(\partial^j / \partial\theta^j)f(\widetilde{X}_i, \theta)$ denote the vectors of partial derivatives with respect to θ of order j of $g(X_i, \theta)$ and $f(\widetilde{X}_i, \theta)$, respectively.

The following assumptions apply to the one-step block GMM, two-step block GMM, or block minimum ρ estimator.

Assumption 1. There is a sequence of iid vectors $\{\varepsilon_i : i = -\infty, ..., \infty\}$ of dimension $L_\varepsilon \geq L_x$ and an $L_x \times 1$ function h such that $X_i = h(\varepsilon_i, \varepsilon_{i-1}, \varepsilon_{i-2}, ...)$. There are constants $K < \infty$ and $\xi > 0$ such that for all $m \geq 1$

$$E||h(\varepsilon_i, \varepsilon_{i-1}, ...) - h(\varepsilon_i, \varepsilon_{i-1}, ..., \varepsilon_{i-m}, 0, 0, ...)|| \leq K \exp(-\xi m).$$

Assumption 2. (a) Θ is compact and θ_0 is an interior point of Θ. (b) Either (i) $\widehat{\theta}_N$ minimizes $J_{N,\pi}(\theta)$ or $J_{N,\pi}(\theta, \widetilde{\theta}_N)$ over $\theta \in \Theta$; θ_0 is the unique solution in Θ to $Eg(X_1, \theta) = 0$; for some function $C_g(x)$, $||g(x, \theta_1) - g(x, \theta_2)|| \leq C_g(x)||\theta_1 - \theta_2||$ for all x in the support of X_1 and all $\theta_1, \theta_2 \in \Theta$; and $EC_g^{q_1}(X_1) < \infty$ and $E||g(X_1, \theta)||^{q_1} < \infty$ for all $\theta \in \Theta$ for all $0 < q_1 < \infty$ or (ii) $\widehat{\theta}_N$ minimizes $N^{-1}\sum_{i=1}^N \rho_\pi(X_i, \theta)$ over $\theta \in \Theta$ for some function $\rho(x, \theta)$ such that $(\partial/\partial\theta)\rho(x, \theta) = g(x, \theta)$ for all x in the support of X_1; θ_0 is the unique minimum of $E\rho(X_1, \theta)$ over $\theta \in \Theta$; and $E\sup_{\theta \in \Theta}||g(X_1, \theta)||^{q_1} < \infty$ and $E|\rho(X_1, \theta)|^{q_1} < \infty$ for all $\theta \in \Theta$ for all $0 < q_1 < \infty$.

Assumption 3. (a) $Eg(X_1, \theta_0)g(X_{1+j}, \theta_0)' = 0$ for all $j > \kappa$ for some $0 \leq \kappa < \infty$. (b) Ω and Ω_0 are positive definite and D is full rank L_θ. (c) $g(x, \theta)$ is $d = d_1 + d_2$ times differentiable with respect to θ on N_0, some neighborhood of θ_0, for all x in the support of X_1, where $d_1 \geq 3$ and $d_2 \geq 0$. (d) There is a function $C_{\partial f}(\widetilde{X}_1)$ such that $||(\partial^j/\partial\theta^j)f(\widetilde{X}_1, \theta) - (\partial^j/\partial\theta^j)f(\widetilde{X}_1, \theta_0)|| \leq C_{\partial f}(\widetilde{X}_1)||\theta - \theta_0||$ for all $\theta \in N_0$ for all $j = 0, ..., d_2$. (e) $EC_{\partial f}^{q_2}(\widetilde{X}_1) < \infty$ and $E||(\partial^j/\partial\theta^j)f(\widetilde{X}_1, \theta_0)||^{q_2} \leq C_f < \infty$ for all $j = 0, ..., d_2$ for some constant C_f (that may depend on q_2) and all $0 < q_2 < \infty$. (f) $f(\widetilde{X}_1, \theta_0)$ is once differentiable with respect to $\widetilde{X}_1$ with uniformly continuous first derivative. (g) If the Wald statistic is considered, the R^{L_η}-valued function $\eta(\cdot)$ is d_1 times continuously differentiable at θ_0 and $(\partial/\partial\theta')\eta(\theta_0)$ is full rank $L_\eta \leq L_\theta$.

Assumption 4. There exist constants $K_1 < \infty$ and $\delta > 0$ such that for arbitrarily large $\zeta > 1$ and all integers $m \in (\delta^{-1}, N)$ and $t \in R^{\dim(f)}$ with $\delta < ||t|| < N^\zeta$,

$$E\left|E\left(\exp\left(it'\sum_{s=1}^{2m+1} f(\widetilde{X}_s, \theta_0)\right)|\{\varepsilon_j : |j - m| > K_1\}\right)\right| \leq \exp(-\delta),$$

where $i = \sqrt{-1}$ here.

The lower bounds on d_1 and d_2 in Assumption 3 are minimal bounds. The results stated below specify more stringent lower bounds that vary depending upon the result.

Assumption 4 is the same as condition (4) of Götze and Hipp (1994). It reduces to the standard Cramér condition if $\{X_i : i \geq 1\}$ are iid. The moment conditions in Assumptions 2 and 3 are stronger than necessary, but lead to relatively simple results. See Andrews (2001a) for a much more complicated set of assumptions, but with weaker moment conditions than those above, that are sufficient for the results given below.

6 Asymptotic Refinements of the Block-block Bootstrap

In this section, we show that the block-block bootstrap leads to greater asymptotic refinements in ERPs of tests and in CI coverage probabilities when compared to the block bootstrap, as well as in comparison to procedures based on first-order asymptotics.

The following Theorem shows that the symmetric two-sided block-block bootstrap t, Wald, and J tests have ERPs of magnitude $o(N^{-(1+\xi)})$ for all $\xi < 1/2 - \gamma$ when the block length ℓ is chosen proportional to N^γ. It shows that the block-block bootstrap equal-tailed two-sided t and one-sided t tests have ERPs of magnitude $o(N^{-(1/2+\xi)})$ for all $\xi < 1/2 - \gamma$ when ℓ is chosen proportional to N^γ. The only restriction on γ is that $0 < \gamma < 1/2$. Hence, for γ close to zero, ξ is close to $1/2$. For m-dependent data, $\gamma = 0$ is permitted.

In contrast, with the block bootstrap, analogous results hold but with the additional restriction that $\xi < \gamma$. The latter restriction, combined with $\xi < 1/2 - \gamma$, imply that $\xi < 1/4$.

The following results hold for statistics based on one-step block GMM, two-step block GMM, and block minimum ρ estimators.

Theorem 1 (a) *Suppose Assumptions* 1–4 *hold with* $d_1 \geq 5$ *and* $d_2 \geq 4$; $0 \leq \xi < 1/2 - \gamma$; $0 < \gamma < 1/2$; $\pi \in (0,1)$; *and* $\pi \to 0$ *and* $\pi\ell - C\log(N) \to \infty$ *as* $N \to \infty$ *for all constants* $0 < C < \infty$. *Then, under* $H_0 : \theta_r = \theta_{0,r}$,

$$P(|T_N| > z^*_{|T|,\alpha}) = \alpha + o(N^{-(1+\xi)}).$$

Under $H_0 : \eta(\theta_0) = 0$,

$$P(\mathcal{W}_N > z^*_{\mathcal{W},\alpha}) = \alpha + o(N^{-(1+\xi)}).$$

In addition, if $L_g > L_\theta$, *then*

$$P(J_N > z^*_{J,\alpha}) = \alpha + o(N^{-(1+\xi)}).$$

(b) *Suppose Assumptions* 1–4 *hold with* $d_1 \geq 4$ *and* $d_2 \geq 3$; $0 \leq \xi < 1/2 - \gamma$; $0 < \gamma < 1/2$; $\pi \in (0,1)$; *and* $\pi \to 0$ *and* $\pi\ell - C\log(N) \to \infty$ *as* $N \to \infty$ *for all* $0 < C < \infty$. *Then, under* $H_0 : \theta_r = \theta_{0,r}$,

$$P(T_N < z^*_{T,\alpha/2} \text{ or } T_N > z^*_{T,1-\alpha/2}) = \alpha + o(N^{-(1/2+\xi)}) \text{ and}$$

$$P(T_N > z^*_{T,\alpha}) = \alpha + o(N^{-(1/2+\xi)}).$$

(c) *If the observations* $\{X_i : i \geq 1\}$ *are* m*-dependent for some integer* $m < \infty$*, then the results of parts* (a) *and* (b) *hold under the stated conditions, but with* $\gamma = 0$ *and with the restrictions on* π *replaced by* $\limsup_{N\to\infty} \pi < 1$ *and* $\liminf_{N\to\infty} \lceil \pi\ell \rceil \geq m+\kappa$.

Comments: 1. The errors in parts (a) and (b) of the Theorem when the critical values are based on standard first-order asymptotics (using the normal distribution or the chi-square distribution) are $O(N^{-1})$, $O(N^{-1/2})$, and $O(N^{-1})$ respectively. Thus, parts (a) and (b) of the Theorem show that the bootstrap critical values reduce the ERP (and the error in CI coverage probability) relative to first-order asymptotics by a factor of at least $N^{-\xi}$. The choice of γ close to zero maximizes ξ subject to the requirement of the Theorem that $\xi < 1/2 - \gamma$. For such a choice of γ, the results of parts (a) and (b) hold for ξ close to 1/2.

2. When the data are m-dependent, part (c) of the Theorem shows that one does not need the block length, ℓ, to diverge to infinity as $N \to \infty$ or the number of observations deleted per block, $\lceil \pi\ell \rceil$, to diverge to infinity as $N \to \infty$. What is needed is that the number of observations deleted per block, $\lceil \pi\ell \rceil$, is greater than to equal to $m + \kappa$ for N large. This suffices, because the blocks statistics are based on sample averages, which are sums of independent blocks provided $\lceil \pi\ell \rceil \geq m + \kappa$, which is exactly mimicked by the independence of the bootstrap blocks.

In contrast, when the block bootstrap is applied to non-block statistics and the observations are m-dependent, the length of the blocks needs to diverge to infinity as $N \to \infty$.

3. The reason that symmetric two-sided block-block bootstrap t tests, Wald tests, and J tests are correct to a higher order than equal-tailed two-sided t tests and one-sided t tests is that the $O(N^{-1/2})$ terms of the Edgeworth expansions of $|T_N|$, $\mathcal{W}_N$, and J_N are zero by a symmetry property. See Hall (1992), Hall and Horowitz (1996), or A2002 for details.

4. The possibility of improving the result of Theorem 1(a) for $|T_N|$ when the data are dependent via the symmetry argument of Hall (1988), which applies with iid data, is unclear, see the discussion in A2002.

7 Monte Carlo Simulations

In this section, we describe some Monte Carlo simulation results that are designed to assess the coverage probability accuracy of block-block bootstrap CIs.

7.1 Experimental Design

We consider a dynamic linear regression model estimated by LS:

$$\begin{aligned} Y_i &= \theta_{0,1} + Y_{i-1}\theta_{0,2} + \sum_{j=3}^{5} Z_{i,j}\theta_{0,j} + U_i \\ &= Z_i'\theta_0 + U_i \text{ for } i = 1, ..., N, \text{ where} \end{aligned}$$

$$
\begin{aligned}
Z_i &= (1, Y_{i-1}, Z_{i,3}, Z_{i,4}, Z_{i,5})', \\
\theta_0 &= (\theta_{0,1}, ..., \theta_{0,5})', \\
Z_{i,j} &= Z_{i-1,j}\rho_Z + V_{i,j} \text{ for } j = 3, 4, 5, \\
X_i &= (Y_i, Z_i')', \text{ and} \\
g(X_i, \theta) &= (Y_i - Z_i'\theta)Z_i. \qquad (7.1)
\end{aligned}
$$

Five regressors are in the model. One is a constant; one is a lagged dependent variable; and the other three are first-order autoregressive (AR(1)) regressors with the same AR(1) parameter ρ_Z. The innovations, $V_{i,j}$, for the AR(1) regressors are iid across i and j with mean zero and variance one and are independent of the errors U_i. The regressor innovations and the errors are taken to have the same distribution. We consider three different distributions: standard normal, chi-square with two degrees of freedom (recentered and rescaled to have mean zero and variance one), and uniform on $[-\sqrt{12}, \sqrt{12}]$ (which has mean zero and variance one). The initial observations used to start up the AR(1) regressors are taken to have the same distribution as the innovations, but are scaled to yield variance stationary processes. The moment vectors $g(X_i, \theta_0)$ are uncorrelated. In terms of the notation introduced above, $\kappa = 0$, $n = N$, and $\widetilde{X}_i = X_i$.

The parameters $\theta_{0,1}, \theta_{0,3}, \theta_{0,4}, \theta_{0,5}$ are taken to be zero. Three combinations of $(\theta_{0,2}, \rho_Z)$ are considered: (.9, .8), (.95, .95), and (.8, .7). Two sample sizes N are considered: 50 and 100.

We consider CIs for the parameter $\theta_{0,2}$ on the lagged dependent variable. The CIs are based on a t statistic that employs the LS estimator of $\theta_{0,2}$ coupled with either a heteroskedasticity consistent standard error estimator or a homoskedasticity consistent standard error estimator:

$$
\begin{aligned}
T_{N,1} &= \frac{N^{1/2}(\widehat{\theta}_{N,2} - \theta_{0,2})}{(\widehat{\sigma}_{N,1})_{22}^{1/2}}, \\
T_{N,2} &= \frac{N^{1/2}(\widehat{\theta}_{N,2} - \theta_{0,2})}{(\widehat{\sigma}_{N,1})_{22}^{1/2}}, \\
\widehat{\theta}_N &= \left(\sum_{i=1}^N Z_i Z_i'\right)^{-1} \sum_{i=1}^N Z_i Y_i, \\
\widehat{\sigma}_{N,1} &= \left(N^{-1}\sum_{i=1}^N Z_i Z_i'\right)^{-1} N^{-1}\sum_{i=1}^N \widehat{U}_i Z_i Z_i' \left(N^{-1}\sum_{i=1}^N Z_i Z_i'\right)^{-1}, \\
\widehat{\sigma}_{N,2} &= N^{-1}\sum_{i=1}^N \widehat{U}_i^2 \left(N^{-1}\sum_{i=1}^N Z_i Z_i'\right)^{-1}, \text{ and} \\
\widehat{U}_i &= Y_i - Z_i'\widehat{\theta}_N. \qquad (7.2)
\end{aligned}
$$

We compare standard two-sided delta method CIs to symmetric two-sided and equal-tailed two-sided block bootstrap and block-block bootstrap CIs. Both non-overlapping and overlapping block bootstrap and block-block bootstrap CIs are considered. The delta method CI is given by $[\widehat{\theta}_{N,2} - z_{\alpha/2}(\widehat{\sigma}_{N,j})_{22}^{1/2}/N^{1/2}, \widehat{\theta}_{N,2} + z_{\alpha/2}(\widehat{\sigma}_{N,j})_{22}^{1/2}/N^{1/2}]$ for $j = 1$ or 2, where $z_{\alpha/2}$ denotes the $1 - \alpha/2$ quantile of the standard normal distribution. The bootstrap CIs are defined in Section 4 above. The bootstrap CIs are based on blocks of length $\ell = 5$ or 10 with the number of observations "skipped" in each block (Skip) in the computation of the block statistics equal to $0, 1$, or 2. Note that the deletion fraction π equals Skip/ℓ. Hence, deletion fractions of 0, .2, and .4 are considered. When Skip = 0, the block-block bootstrap reduces to the standard block bootstrap.

The number of simulation repetitions used is 40,000 for each case considered. This yields simulation standard errors of (approximately) .0010, .0015, and .0004 for the simulated coverage probabilities of nominal 95%, 90%, and 99% CIs respectively.

7.2 Simulation Results

Table I reports results for *symmetric* two-sided CIs. Table II does likewise for *equal-tailed* two-sided CIs. In both Tables, results for a base case are reported in Column 1 and variations on the base case are reported in Columns 2-9. The base case has $(\theta_{0,2}, \rho_Z) = (.9, .8)$, standard normal (N(0, 1)) distributions for the regressor innovations and errors, sample size $N = 50$, CIs based on heteroskedasticity consistent (HC) standard error estimates, and CIs with 95% nominal coverage probability. In Columns 2 and 3, the nominal coverage probabilities are 90% and 99%, respectively, and all other features are as in the base case. Columns 4 and 5 differ from the base case in that $(\theta_{0,2}, \rho_Z) = (.95, .95)$ and $(.8, .7)$ respectively. Column 6 differs from the base case in that $N = 100$. Columns 7 and 8 differ from the base case in that the distributions of the regressor innovations and errors are chi-square with two degrees of freedom (χ_2^2) and uniform on $[-\sqrt{12}, \sqrt{12}]$ (Unif) respectively. Column 9 differs from the base case in that homoskedastic (Homo) standard error estimates are employed.

The results of Table I show the following:

1. The coverage probabilities of the delta method CIs are poor. For example, in the base case, the coverage probability of the nominal 95% delta method CI is .759.

2. The bootstrap CIs perform fairly well and, hence, out-perform the delta method CIs by a wide margin. This is true regardless of the choice of the block length ℓ, the number of observations skipped, and the case considered. For example, in the base case, the worst nominal 95% bootstrap CI has coverage probability .915. The best has coverage probability .942.

3. The best bootstrap results are quite good and are obtained with (ℓ, Skip) = $(10, 2)$ or $(5, 1)$ with non-overlapping or overlapping blocks. These bootstraps are based on block statistics that skip 20% of the observations (which yields standard errors estimates that are 12% larger than when Skip = 0). For example, for (ℓ, Skip) = $(10, 2)$, the non-overlapping block-block bootstrap CI has coverage probability that varies between .928 and .955 for the seven cases with 95% nominal CIs.

4. When Skip $= 0$, the bootstrap results are not very sensitive to the choice of block length ℓ. When Skip ≥ 1, the bootstrap results are not very sensitive to the block length provided Skip is adjusted so that the same deletion fraction π is maintained. That is, the results for (ℓ, Skip) $= (10, 2)$ are quite similar to those with (ℓ, Skip) $= (5, 1)$.

5. The coverage probabilities of the bootstrap CIs are increasing in the number of observations skipped in almost all cases. Because the standard block bootstrap (for which Skip $= 0$) under-covers in all cases, this leads to smaller coverage probability errors in most cases for one or more bootstraps based on block statistics (for which Skip ≥ 1).

6. There is not much difference in performance between the non-overlapping and the overlapping bootstraps. Neither is dominant.

7. The results for nominal 90% and 99% CIs given in Columns 2 and 3 are analogous to those in Column 1 for 95% CIs. That is, the delta method CI does quite poorly and the bootstrap CIs do fairly well.

8. The effect of increasing and decreasing the amount of correlation, as shown in Columns 4 and 5, respectively, is as expected. Increasing the amount of correlation reduces the coverage probabilities and increases the coverage probability errors of all CIs. Decreasing the amount of correlation increases the coverage probabilities and decreases the coverage probability errors.

9. The effect of increasing the sample size, as shown in Column 6, is to increase the coverage probabilities and reduce the coverage probability errors for almost all CIs.

10. The effect of changing from normal to χ^2_2 and uniform distributions is to increase and decrease, respectively, the coverage probabilities of the CIs. Thus, the delta method CI and most bootstrap CIs perform better with χ^2_2 distributions and worse with uniform distributions than with standard normal distributions.

11. The use of the homoskedasticity consistent standard error estimate (which is a consistent estimate in the cases considered) improves the coverage probability of the delta method CI, but has little effect on the bootstrap CIs. Hence, the bootstrap CIs sacrifice little by using heteroskedasticity consistent standard error estimates.

For brevity, Table II reports results only for non-overlapping bootstrap CIs. Analogous results for overlapping bootstraps are similar, though slightly worse. Table II reports three probabilities for each CI—the coverage probability, the probability of missing to the left, and the probability of missing to the right.

The results of Table II show the following:

1. The coverage probabilities of the equal-tailed bootstrap CIs are noticeably lower than those of the symmetric bootstrap CIs reported in Table I. For example, in

the base case, the coverage probability of the nominal 95% equal-tailed bootstrap CI with (ℓ, Skip) $= (10, 2)$ is .885, whereas that of the corresponding symmetric CI is .942.

2. The coverage probabilities of the equal-tailed bootstrap CIs are noticeably better than those of the delta method.

3. The best bootstrap CI is the one with (ℓ, Skip) $= (5, 2)$. The block LS estimator upon which this CI is based has standard errors that are 29% larger than those of the full sample LS estimator. In consequence, this bootstrap CIs is noticeably longer than those based on the standard block bootstrap (for which Skip $= 0$).

4. The probability of missing to the left is too high for all CIs, but especially for the delta method CI. This reflects the downward bias of the LS estimator.

5. The probability of the delta method CI missing to the right is too low. For the bootstrap CIs, it is sometimes too high and sometimes too low.

6. The effect of changing the amount of correlation, sample size, distribution, and standard error estimate is similar for equal-tailed CIs as for symmetric CIs.

In sum, the Monte Carlo results show that all of the bootstrap CIs considered out-perform the delta method CIs. The margin of improvement is quite substantial for symmetric bootstrap CIs. The results also show that the block-block bootstrap yields improved coverage probabilities in the majority of cases considered compared to the standard block bootstrap. Hence, there is some evidence that the theoretical advantages established above for the block-block bootstrap are reflected in finite samples.

8 Appendix of Proofs

The proof of Theorem 1 holds by making some adjustments to the proof of Theorem 2 of A2002. The proof of Theorem 2 of A2002 relies on sixteen Lemmas. These Lemmas need to be adjusted as follows. Lemma 1 needs to hold for triangular arrays of functions $\{h_{N,i}(\cdot) : i \leq N, N \geq 1\}$, rather than a single function $h(\cdot)$, in order to apply the Lemma with $h_{N,i}(X_i) = g_\pi(X_i, \theta_0)$, rather than $h(X_i) = g(X_i, \theta_0)$. This extension is easily achieved. It is stated as Lemma 1 below.

Given the new Lemma 1 (and the fact that $(N\tau)/N \to 1$ as $N \to \infty$ under the assumption that $\pi = 1 - \tau \to 0$), the proofs of Lemmas 2–13 and 16 of A2002 hold with $g(X_i, \theta_0)$ replaced by $g_\pi(X_i, \theta_0)$ throughout without any significant changes in their proofs. Lemma 15 of A2002 is not needed when block statistics are considered because it involves the behavior of correction factors, which are not used with block statistics. Lemma 14 of A2002 needs to be changed. In particular, we need to show that it holds with the condition $\xi < \gamma$ deleted. Lemma 2 below gives the required result.

Given that Lemmas 2–14 and 16 of A2002 hold with $g(X_i, \theta_0)$ replaced by $g_\pi(X_i, \theta_0)$, Theorem 1(a) and (b) hold by the proof of Theorem 2 of A2002. For the case of m-dependent observations $\{X_i : i \geq 1\}$ (covered in Theorem 1(c)), the only adjustment to the proof that is required is that the result of Lemma 14 of A2002 needs to hold with $\gamma = 0$. Lemma 2 below covers this case.

8.1 Lemmas

Lemma 1 *Suppose Assumption* 1 *holds.*
(a) *Let* $\{h_{N,i}(\cdot) : i \leq N, N \geq 1\}$ *be a triangular array of matrix-valued functions that satisfy* $Eh_{N,i}(\widetilde{X}_i) = 0$ *for all* i, N *and* $\sup_{i \leq N, N \geq 1} E\|h_{N,i}(\widetilde{X}_i)\|^p < \infty$ *for* $p \geq 2$ *and* $p > 2a/(1-2c)$ *for some* $c \in [0, 1/2)$ *and* $a \geq 0$. *Then, for all* $\varepsilon > 0$,

$$\lim_{N\to\infty} N^a P(\|N^{-1}\sum_{i=1}^N h_{N,i}(\widetilde{X}_i)\| > N^{-c}\varepsilon) = 0.$$

(b) *Let* $\{h_{N,i}(\cdot) : i \leq N, N \geq 1\}$ *be a triangular array of matrix-valued functions that satisfy* $\sup_{i \leq N, N \geq 1} E\|h(\widetilde{X}_i)\|^p < \infty$ *for* $p \geq 2$ *and* $p > 2a$ *for some* $a \geq 0$. *Then, there exists a constant* $K < \infty$ *such that*

$$\lim_{N\to\infty} N^a P(\|N^{-1}\sum_{i=1}^N h_{N,i}(\widetilde{X}_i)\| > K) = 0.$$

Asymptotic refinements of the block bootstrap depend on the differences between the Edgeworth expansions of the dfs of T_N and T_N^* being small (and analogously for ($\mathcal{W}_N$, $\mathcal{W}_N^*$) and (J_N, J_N^*)). Let $\nu_{N,\pi,a}$ denote a vector of population moments including those of $g_\pi(X_i, \theta_0)$ and some of its partial derivatives with respect to θ. $\nu_{N,\pi,a}$ is defined precisely below. Let $\nu^*_{N,\pi,a}$ denote an analogous vector of bootstrap moments including those of $g^*_\pi(X^*_i, \theta_0)$ and some of its partial derivatives. Edgeworth expansions of the dfs of T_N, $\mathcal{W}_N$, and J_N at a point y, with remainder of order $o(N^{-a})$, where $2a$ is an integer, depend on polynomials in y whose coefficients are polynomials in the elements of $\nu_{N,\pi,a}$. Analogously, Edgeworth expansions of the dfs of T_N^*, $\mathcal{W}_N^*$, and J_N^* are the same as those of T_N, $\mathcal{W}_N$, and J_N, but with $\nu_{N,\pi,a}^*$ in place of $\nu_{N,\pi,a}$. In consequence, asymptotic refinements of the block bootstrap depend on the magnitude of the differences between $\nu_{N,\pi,a}^*$ and $\nu_{N,\pi,a}$. Lemma 2 shows that these differences are small asymptotically.

We now define $\nu_{N,\pi,a}$ and $\nu_{N,\pi,a}^*$ precisely. Let $f(\widetilde{X}_i,\theta)$ be the vector-valued function defined at the beginning of Section 5. Let $f_\pi(\widetilde{X}_i,\theta)$ be the function derived from $f(\widetilde{X}_i,\theta)$ in the same way as $g_\pi(\widetilde{X}_i,\theta)$ is derived from $g(\widetilde{X}_i,\theta)$ in (3.3). Let $f_\pi^*(\widetilde{X}_i^*,\theta)$ denote the vector containing the unique components of $g_\pi^*(X_i^*,\theta)$ and $g_\pi^*(X_i^*,\theta)g^*(X_{i+j}^*,\theta)'$ for all $j=0,...,\kappa$ and their derivatives with respect to θ through order d_1. Let $S_{N,\pi}=(N\tau)^{-1}\sum_{i=1}^N f_\pi(\widetilde{X}_i,\theta_0)$, $S_\pi=ES_{N,\pi}$, $S_{N,\pi}^*=(N\tau)^{-1}\sum_{i=1}^N f_\pi^*(\widetilde{X}_i^*,\widehat{\theta}_N)$, and $S_\pi^*=E^*S_{N,\pi}^*$. Let $\Psi_{N,\pi}=(N\tau)^{1/2}(S_{N,\pi}-S_\pi)$ and $\Psi_{N,\pi}^*=(N\tau)^{1/2}(S_{N,\pi}^*-S_\pi^*)$. Let $\Psi_{N,\pi,j}$ and $\Psi_{N,\pi,j}^*$ denote the j-th elements of $\Psi_{N,\pi}$ and $\Psi_{N,\pi}^*$, respectively. Let $\nu_{N,\pi,,a}$ and $\nu_{N,\pi,a}^*$ denote vectors of moments of the form $(N\tau)^{\alpha(m)}E\prod_{\mu=1}^m\Psi_{N,\pi,j_\mu}$ and $(N\tau)^{\alpha(m)}\ E^*\prod_{\mu=1}^m\Psi_{N,\pi,j_\mu}^*$, respectively, where $2\leq m\leq 2a+2$, $\alpha(m)=0$ if m is even, and $\alpha(m)=1/2$ if m is odd.

Lemma 2 *Suppose Assumptions* 1 *and* 3 *hold with* $d_2\geq 2a+1$ *for some* $a\geq 0$, $0\leq\xi<1/2-\gamma$, *and either* (i) $0<\gamma<1/2$ *or* (ii) *the observations* $\{X_i : i\geq 1\}$ *are* m*-dependent for some integer* $m<\infty$, $\gamma=0$, $\pi\in(0,1)$, *and* $\liminf_{N\to\infty}\lceil\pi\ell\rceil\geq m+\kappa$. *Then,*

$$\lim_{N\to\infty}N^aP(||\nu_{N,\pi,a}^*-\nu_{N,\pi,a}||>(N\tau)^{-\xi})=0.$$

Comment. The condition $\xi<\gamma$, which is needed in Lemma 14 of A2002, is not needed in Lemma 2 because the moments considered are moments of block statistics. This is the key feature of block statistics that allows the block-block bootstrap to attain larger asymptotic refinements than the block bootstrap applied to standard statistics.

8.2 Proofs of Lemmas

8.2.1 Proof of Lemma 1

A strong mixing moment inequality of Yokoyama (1980) and Doukhan (1995, Theorem 2 and Remark 2, pp. 25–30) gives $E||\sum_{i=1}^N h_{N,i}(\widetilde{X}_i)||^p<CN^{p/2}$ provided $p\geq 2$, where C does not depend on N. Application of Markov's inequality and the Yokoyama–Doukhan inequality yields the left-hand side in part (a) of the Lemma to be less than or equal to

$$\lim_{N\to\infty}\varepsilon^{-p}N^{a-p+pc}E||\sum_{i=1}^N h_{N,i}(\widetilde{X}_i)||^p\leq\lim_{N\to\infty}\varepsilon^{-p}CN^{a-p+pc+p/2}=0. \tag{8.1}$$

Part (b) follows from part (a) applied to $h_{N,i}(\widetilde{X}_i)-Eh_{N,i}(\widetilde{X}_i)$ with $c=0$ and the triangle inequality. $\square$

8.2.2 Proof of Lemma 2

The proof of Lemma 14 of A2002 goes through with $g(X_i, \theta_0)$ replaced by $g_\pi(X_i, \theta_0)$ except for the proof that $B_2 = 0$.

More specifically, as in A2002, the least favorable value of m for the bootstrap moment $(N\tau)^{\alpha(m)} E^* \prod_{\mu=1}^{m} \Psi^*_{N,\pi,j_\mu}$ (in terms of its distance from the corresponding population moment) is three. Hence, we just consider this case. For notational simplicity, suppose $j_\mu = 1$ for $\mu = 1, 2, 3$. Thus, we need to show that

$$\lim_{N\to\infty} N^a P(|(N\tau)^{1/2} E^*(\Psi^*_{N,\pi,1})^3 - (N\tau)^{1/2} E\Psi^3_{N,\pi,1})| > (N\tau)^{-\xi}) = 0. \tag{8.2}$$

Let $f_{\pi,i} = f_{\pi,1}(\widetilde{X}_i, \theta_0) - Ef_{\pi,1}(\widetilde{X}_i, \theta_0)$, where $f_{\pi,1}(\widetilde{X}_i, \theta_0)$ denotes the first element of $f_\pi(\widetilde{X}_i, \theta_0)$. Let $b_1 = \{1, ..., \ell\}$, $b_2 = \{\ell+1, ..., 2\ell\}$, ..., $b_b = \{(b-1)\ell+1, ..., b\ell\}$, where $N = b\ell$. Let $Y_{\pi,j} = \sum_{i\in b_j} f_{\pi,i}$. Then,

$$\Psi_{N,\pi,1} = (N\tau)^{-1/2} \sum_{i=1}^{N} f_{\pi,i} = (N\tau)^{-1/2} \sum_{j=1}^{b} Y_{\pi,j}. \tag{8.3}$$

By the arguments in the proof of Lemma 14 of A2002, provided $\xi < 1/2 - \gamma$,

$$\lim_{N\to\infty} N^a P(|(N\tau)^{1/2} E^*(\Psi^*_{N,\pi,1})^3 - (N\tau)^{-1} bEY^3_{\pi,1})| > (N\tau)^{-\xi}) = 0. \tag{8.4}$$

Hence, it suffices to show that

$$\limsup_{N\to\infty} (N\tau)^{\xi} |(N\tau)^{1/2} E\Psi^3_{N,\pi,1} - (N\tau)^{-1} bEY^3_{\pi,1}| = 0. \tag{8.5}$$

(Equation (8.5) shows that B_2 of A2002 equals zero.)

Using (8.3), we have

$$(N\tau)^{1/2} E\Psi^3_{N,\pi,1} = (N\tau)^{\xi-1} \sum_{j_1=1}^{b} \sum_{j_2=1}^{b} \sum_{j_3=1}^{b} EY_{j_1} Y_{j_2} Y_{j_3}. \tag{8.6}$$

Hence,

$$(N\tau)^{\xi} |(N\tau)^{1/2} E\Psi^3_{N,\pi,1} - (N\tau)^{-1} bEY^3_{\pi,1}| = (N\tau)^{\xi-1} \sum_{\substack{j_1=1 \\ j_1\neq j_2\neq j_3}}^{b} \sum_{j_2=1}^{b} \sum_{j_3=1}^{b} EY_{j_1} Y_{j_2} Y_{j_3}. \tag{8.7}$$

If the observations $\{X_i : i \geq 1\}$ are m-dependent, then the observations $\{\widetilde{X}_i : i \geq 1\}$ are $(m + \kappa)$-dependent and Y_{j_1} and Y_{j_2} are independent for all $j_1 \neq j_2$ for N large because the number of deleted observations at the end of each block satisfies $\lceil \pi\ell \rceil \geq m + \kappa$ for N large. Since $EY_{j_1} = 0$ for all j_1, the rhs of (8.7) equals zero in the m-dependent case.

Next, we consider the case where the observations are not necessarily m-dependent, but $0 < \gamma < 1/2$. By a strong mixing covariance inequality of Davydov, e.g., see Doukhan (1995, Thm. 3(1), p. 9),

$$|EY_{j_1}Y_{j_2}Y_{j_3}| \leq 8||Y_{j_1}||_p \cdot ||Y_{j_2}Y_{j_3}||_q \alpha^r(\lceil \pi\ell \rceil - \kappa), \tag{8.8}$$

where $p, q, r \geq 1$, $1/p + 1/q + 1/r = 1$, $|| \cdot ||_p$ denotes the L^p norm, and $\{\alpha(s) : s \geq 1\}$ are the strong mixing numbers of $\{X_i : i \geq 1\}$, which decline to zero exponentially fast by Assumption 1. This inequality holds because the summands in Y_{j_1} are separated from those in any blocks Y_{j_2} and Y_{j_3} by at least $\lceil \pi\ell \rceil - \kappa$ by the block feature of $f_{\pi,i}$. This is the key part of the proof.

Next, by Minkowski's inequality, $||Y_{j_1}||_p = ||Y_1||_p \leq \ell||f_{\pi,1}||_p \leq \ell\tau C_1$ for some constant $C_1 < \infty$. By an application of the Cauchy-Swartz inequality and the fact that $||Y_2||_{2q} = ||Y_3||_{2q}$, we have $||Y_{j_2}Y_{j_3}||_q = ||Y_2Y_3||_q \leq ||Y_2||_{2q}^2 \leq (\ell\tau C_1)^2$. Hence, (8.5) holds by (8.7) and (8.8) provided

$$(N\tau)^{\xi-1}b^3(\ell\tau)^3\alpha^r(\lceil \pi\ell \rceil - \kappa) = (N\tau)^{2+\xi}\alpha^r(\lceil \pi\ell \rceil - \kappa) \to 0 \text{ as } N \to \infty. \tag{8.9}$$

Since the α–mixing numbers, $\{\alpha(s) : s \geq 1\}$, decline to zero exponentially fast in s, (8.9) holds provided $\pi\ell - C\log(N) \to \infty$ for all $C < \infty$, as is assumed. $\square$

Footnotes

[1] The author thanks Joel Horowitz for helpful comments. The author gratefully acknowledges the research support of the National Science Foundation via grant number SBR-9730277.

[2] These correction factors alleviate, but do not solve, the join-point problem for the block bootstrap applied to standard statistics. They allow the block bootstrap to attain ERPs for two-sided t tests of magnitude $O(N^{-1-\xi})$ for $\xi < 1/4$, but these are still noticeably larger than those attained by the nonparametric iid bootstrap. Correction factors are not needed for one-sided t tests to yield asymptotic refinements.

[3] Furthermore, one could consider block statistics that are defined using a smooth tapering function. The block bootstrap applied to such statistics would be a tapered bootstrap. It is likely that the block bootstrap applied to such statistics would yield asymptotic refinements akin to those obtained in this paper.

[4] The condition $\xi < 1/2 - \gamma$ is stronger than necessary to establish the stated result, which applies to a bootstrap moment of order two. However, this condition is necessary when bootstrap moments of order three, such as $E^*(N^{-1/2}\sum_{i=1}^{N}(\partial/\partial\theta_1)\rho(X_i^*, \theta_0))^3$, are considered.

[5] For convenience, we state that limits are as $N \to \infty$ below, although, strictly speaking, they are limits as $n \to \infty$.

References

Andrews, D. W. K. (2001a): "Higher-order Improvements of a Computationally Attractive k-step Bootstrap for Extremum Estimators," Cowles Foundation Discussion Paper No. 1230R, Yale University. Available at http://cowles.econ.yale .edu.

————(2001b): "Higher-order Improvements of the Parametric Bootstrap for Markov Processes," Cowles Foundation Discussion Paper No. 1334, Yale University. Available at http://cowles.econ.yale.edu.

————(2002): "Higher-order Improvements of a Computationally Attractive k-step Bootstrap for Extremum Estimators," *Econometrica,* 70, 119–162.

Andrews, D. W. K. and M. Buchinsky (2000): "A Three-step Method for Choosing the Number of Bootstrap Repetitions," *Econometrica*, 68, 23–51.

Bhattacharya, R. N. (1987): "Some Aspects of Edgeworth Expansions in Statistics and Probability," in *New Perspectives in Theoretical and Applied Statistics*, ed. by M. L. Puri, J. P. Vilaploma, and W. Wertz. New York: Wiley, 157–170.

Bhattacharya, R. N. and J. K. Ghosh (1978): "On the Validity of the Formal Edgeworth Expansion," *Annals of Statistics*, 6, 434–451.

Brown, B. W. and S. Maital (1981): "What Do Economists Know? An Empirical Study of Experts' Expectations," *Econometrica,* 49, 491–504.

Carlstein, E. (1986): "The Use of Subseries Methods for Estimating the Variance of a General Statistic from a Stationary Time Series," *Annals of Statistics*, 14, 1171–1179.

Chandra, T. K. and J. K. Ghosh (1979): "Valid Asymptotic Expansions for the Likelihood Ratio Statistic and Other Perturbed Chi-square Variables," *Sankhya*, 41, Series A, 22–47.

Doukhan, P. (1995): *Mixing: Properties and Examples.* New York: Springer-Verlag.

Götze, F. and C. Hipp (1983): "Asymptotic Expansions for Sums of Weakly Dependent Random Vectors," *Z. Wahrscheinlichkeitstheorie verw. Gebiete*, 64, 211–239.

————(1994): "Asymptotic Distribution of Statistics in Time Series," *Annals of Statistics*, 22, 2062–2088.

Götze, F. and H. R. Künsch (1996): "Second-order Correctness of the Blockwise Bootstrap for Stationary Observations," *Annals of Statistics,* 24, 1914-1933.

Hall, P. (1988): "On Symmetric Bootstrap Confidence Intervals," *Journal of the Royal Statistical Society,* Series B, 50, 35–45.

———(1992): *The Bootstrap and Edgeworth Expansion.* New York: Springer–Verlag.

Hall, P. and J. L. Horowitz (1996): "Bootstrap Critical Values for Tests Based on Generalized-Method-of-Moment Estimators," *Econometrica,* 64, 891–916.

Hansen, L. P. and R. J. Hodrick (1980): "Forward Exchange Rates as Optimal Predictors of Future Spot Rates: An Econometric Analysis," *Journal of Political Economy,* 88, 829–853.

Hansen, L. P. and K. J. Singleton (1982): "Generalized Instrumental Variables Estimation of Nonlinear Rational Expectations Models," *Econometrica,* 50, 1269–1286.

———(1996): "Efficient Estimation of Linear Asset-Pricing Models with Moving Average Errors," *Journal of Business and Economic Statistics,* 14,53–68.

Horowitz, J. L. (2001): "Bootstrap Methods for Markov Processes," unpublished manuscript, Department of Economics, Northwestern University.

Inoue, A. and M. Shintani (2000): "Bootstrapping GMM Estimators for Time Series," unpublished working paper, Department of Economics, Vanderbilt University.

Künsch, H. R. (1989): "The Jackknife and the Bootstrap for General Stationary Observations," *Annals of Statistics*, 17, 1217–1241.

Lahiri, S. N. (1996): "On Edgeworth Expansion and Moving Block Bootstrap for Studentized M-estimators in Multiple Linear Regression Models," *Journal of Multivariate Analysis,* 56, 42-59.

McCallum, B. T. (1979): "Topics Concerning the Formulation, Estimation, and Use of Macroeconometric Models with Rational Expectations," *American Statistical Association Proceedings of the Business and Ecnomic Statistics Section,* 65–72.

Paparoditis, E. and Politis, D. N. (2001): "Tapered Block Bootstrap," *Biometrika,* 88, 1105–1119.

——— (2002): "The Tapered Block Bootstrap for General Statistics from Stationary Sequences," *Econometrics Journal,* forthcoming.

Shorack, G. R. (1982): "Bootstrapping Robust Regression," *Communications in Statistics–Theory and Methods,* 11, 961–972.

Yokoyama, R. (1980): "Moment Bounds for Stationary Mixing Sequences," *Z. Wahrscheinlichkeitstheorie verw. Gebiete*, 52, 45–57.

Zvingelis, Y. (2002): "On Bootstrap Coverage Probability with Dependent Data," forthcoming in *Computer-aided Econometrics,* ed. by D. Giles. New York: Marcel Dekker.

TABLE I

Coverage Probabilities for Two-sided Delta Method and Symmetric Two-sided Block and Block-Block Bootstrap Confidence Interals*

Column		1	2	3	4	5	6	7	8	9
$\theta_{0,2}$		**.9**	.9	.9	**.95**	**.8**	.9	.9	.9	.9
ρ_Z		**.8**	.8	.8	**.95**	**.7**	.8	.8	.8	.8
N		**50**	50	50	50	50	**100**	50	50	50
Distribution		**N(0,1)**	N(0,1)	N(0,1)	N(0,1)	N(0,1)	N(0,1)	$\boldsymbol{\chi^2_2}$	**Unif**	N(0,1)
Std Err Est		**HC**	HC	HC	HC	HC	HC	HC	HC	**Homo**
Conf Level		**95%**	**90%**	**99%**	95%	95%	95%	95%	95%	95%
Δ Method		.759	.669	.886	.702	.833	.853	.791	.751	.808
Non-overlapping Block Bootstrap										
ℓ	Skip									
5	0	.922	.870	.976	.903	.945	.941	.942	.920	.921
5	1	.928	.876	.982	.916	.947	.935	.948	.926	.926
5	2	.966	.932	.994	.959	.975	.966	.975	.961	.966
10	0	.915	.870	.972	.903	.935	.939	.931	.913	.914
10	1	.920	.873	.977	.908	.936	.939	.935	.919	.917
10	2	.938	.900	.984	.928	.952	.955	.952	.938	.937
Overlapping Block Bootstrap										
ℓ	Skip									
5	0	.925	.875	.977	.904	.949	.945	.947	.921	.924
5	1	.939	.895	.983	.923	.958	.953	.962	.933	.938
5	2	.958	.923	.989	.947	.969	.962	.974	.951	.958
10	0	.923	.878	.975	.910	.943	.945	.940	.923	.923
10	1	.930	.889	.979	.919	.948	.948	.950	.930	.930
10	2	.942	.902	.985	.931	.956	.956	.959	.941	.941

* The standard block bootstrap results are those for which Skip = 0 and the block-block bootstrap results are those for which Skip = 1 or 2.

TABLE II

Coverage Probabilities and Probabilities of CIs Missing to the Left and Right for Two-sided Delta Method and Two-sided Equal-tailed Block and Block-Block Bootstrap Confidence Interals: Non-overlapping Blocks

Column		1	2	3	4	5	6	7	8	9
$\theta_{0,2}$		**.9**	.9	.9	**.95**	**.8**	.9	.9	.9	.9
ρ_Z		**.8**	.8	.8	**.95**	**.7**	.8	.8	.8	.8
N		**50**	50	50	50	50	**100**	50	50	50
Distribution		**N(0,1)**	N(0,1)	N(0,1)	N(0,1)	N(0,1)	N(0,1)	$\boldsymbol{\chi^2_2}$	**Unif**	N(0,1)
Std Err Est		**HC**	HC	HC	HC	HC	HC	HC	HC	**Homo**
Conf Level		**95%**	**90%**	**99%**	95%	95%	95%	95%	95%	95%
A. Coverage Probabilities										
Δ Method		.759	.669	.887	.702	.833	.853	.791	.751	.808
ℓ	Skip									
5	0	.848	.764	.946	.825	.876	.874	.840	.863	.848
5	1	.890	.816	.967	.875	.912	.906	.883	.901	.891
5	2	.935	.879	.986	.926	.950	.940	.931	.939	.939
10	0	.840	.770	.936	.812	.851	.884	.829	.851	.841
10	1	.861	.796	.950	.835	.873	.901	.853	.873	.861
10	2	.885	.826	.960	.862	.895	.920	.882	.896	.887
B. Probabilities of Missing to the Left—Nominal Probabilities are .025, .05, or .005										
Δ Method		.231	.314	.111	.280	.153	.138	.200	.239	.185
ℓ	Skip									
5	0	.137	.211	.049	.149	.098	.111	.143	.123	.137
5	1	.100	.165	.030	.107	.071	.084	.104	.090	.099
5	2	.059	.110	.013	.064	.041	.054	.062	.057	.057
10	0	.133	.191	.050	.149	.104	.095	.143	.123	.131
10	1	.116	.168	.039	.130	.088	.081	.121	.105	.115
10	2	.095	.143	.031	.107	.072	.066	.097	.086	.094
C. Probabilities of Missing to the Right—Nominal Probabilities are .025, .05, or .005										
Δ Method		.009	.017	.003	.018	.015	.009	.009	.010	.006
ℓ	Skip									
5	0	.015	.025	.005	.025	.026	.015	.017	.013	.015
5	1	.010	.019	.003	.018	.017	.010	.013	.009	.010
5	2	.006	.011	.001	.011	.009	.005	.007	.005	.004
10	0	.027	.040	.014	.039	.045	.021	.029	.026	.027
10	1	.024	.036	.011	.036	.039	.018	.026	.023	.024
10	2	.020	.031	.009	.031	.033	.014	.022	.018	.020